Spatial and Temporal Statistical Analysis of a Ground-Water Level Network, Broward County, Florida

By Eric D. Swain and Roy S. Sonenshein

U.S. GEOLOGICAL SURVEY
Water-Resources Investigations Report 94-4076

Prepared in cooperation with the
SOUTH FLORIDA WATER MANAGEMENT DISTRICT

Tallahassee, Florida
1994

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
GORDON P. EATON, Director

For additional information, write to:

District Chief
U.S. Geological Survey
Suite 3015
227 N. Bronough Street
Tallahassee, FL 32301

Copies of this report can be purchased from:

U.S. Geological Survey
Earth Science Information Center
Open-File Reports Section
Box 25286, MS 517
Denver Federal Center
Denver, CO 80225

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS AND ACRONYMS

Multiply	By	To obtain
inch (in.)	25.4	millimeter
foot (ft)	0.3048	meter
mile (mi)	1.609	kilometer
foot squared (ft^2)	0.0w9294	square meter
square mile (mi^2)	2.590	square kilometer
foot per day (ft/d)	0.3048	meter per day

Acronyms

CP	Confidence Polygon
GIS	Geographic Information System
STARE	Spatial and Temporal Adequacy and Redundancy Evaluation
SFWMD	South Florida Water Management District
USGS	U.S. Geological Survey

Spatial and Temporal Statistical Analysis of a Ground-Water Level Network, Broward County, Florida

By Eric D. Swain and Roy S. Sonenshein

Abstract

The U.S. Geological Survey has developed a method to evaluate the spatial and temporal statistics of a continuous ground-water level recorder network in Broward County, Florida. Because the Broward County network is sparse for most spatial statistics, a technique has been developed to define polygons for each well that represent the area monitored by the well within specified criteria. The boundaries of these "confidence polygons" are defined by the endpoints of radial lines oriented toward the other wells. The lengths of these lines are determined as the statistically estimated distances to the points at which ground-water levels can be predicted within specified criteria. The confidence polygons indicate: (1) the areal coverage of the network, (2) locations where data are unavailable, and (3) areas of redundant data collection. Comparison with data from a non-continuous recorder well indicates that the confidence polygons are a good representation of areal coverages.

The temporal analysis utilizes statistical techniques similar to those used in the spatial method, defining variations in time rather than in space. Consequently, instead of defining radial distances to points, time intervals are defined over which water-level values can be predicted within a specified confidence. These "temporal confidence intervals" correspond to maximum allowable periods between field measurements.

To combine all results from the analyses, a single coefficient reflecting the spatial and temporal results has been developed. The coefficient is referred to as the Spatial and Temporal Adequacy and Redundancy Evaluation (STARE) and is determined by three factors: the size of the confidence polygon, the number of times the well is part of a redundant pair, and the temporal confidence interval. This coefficient and the individual results of each analysis are used in evaluating the present network and determining future management decisions.

INTRODUCTION

Networks for monitoring saltwater intrusion, ground-water quality, and water-table elevations have been developed in Broward County, Fla., but these networks have not kept up with the growth in the county. Other networks, such as those set up by well-field operators for well-field protection, have not been included in a comprehensive ground-water monitoring system.

The U.S. Geological Survey (USGS) began development of a ground-water level recorder network in Broward County, beginning in 1940, and at present (1994) continuously monitors 36 wells. As the population of Broward County grew and the urbanized area developed, new wells were installed. Some of the new wells were installed to develop the network; other wells were preexisting and were used for some other purpose, such as well-field analysis. New canals and water-management structures, changing land use, and expansion of municipal well fields have resulted in wells no longer monitoring the situations for which they were originally designed. A need exists to evaluate the current ground-water monitoring network, develop criteria for future monitoring needs, and design optimal regional monitoring networks.

The sample size of water-level measurements in the Broward County ground-water level monitoring network is large enough, both in spatial distribution of the wells and in temporal length of record, to perform statistical analyses that describe the spatial and temporal attributes of the water-level variations in the study area (fig. 1). A spatial statistical analysis of this water-level monitoring network can compare each well with every other well in the network to determine relations between water levels. Thus, water levels in wells that are similar to those in adjacent wells can be considered redundant. Areas between wells with unrelated water levels are possible locations for new wells. A method for determining which wells are redundant and locations where water-level data are lacking could facilitate future planning in ground-water network design.

The frequency of well measurements is also important in ground-water level monitoring. An inappropriate measurement interval might result in unnecessary expenditures or insufficient data. The present measurement scheme in the Broward County water-level monitoring network considers no temporal properties of the water levels in determining measurement intervals. A statistical analysis of the temporal variation in water levels at continuous recorder stations can be used to indicate time periods over which water levels are similar. This information can be used to determine more efficient measurement schedules for the 36 continuous recording wells located in the Broward County ground-water-level monitoring network (fig. 2). Consideration of these factors is important in designing an efficient and effective measurement scheme.

The USGS, in cooperation with the South Florida Water Management District (SFWMD), conducted a study to: (1) develop methods for designing monitoring networks based on mathematical models and statistical techniques, (2) examine regional monitoring networks with the ultimate objective of eliminating existing monitoring wells that are redundant, (3) determine locations where additional wells are needed, and (4) optimize temporal measurements.

Purpose and Scope

This report documents statistical techniques developed for an analysis of the: (1) spatial coverage of a well network, (2) redundancy of a well network, and (3) optimal water-level measurement intervals for a ground-water level monitoring network. Because the Broward County water-level well network is sparse, the new method defines the spatial coverage of wells in terms of overlapping polygons that indicate the degree of correlation with adjacent wells. This method more accurately depicts the sparseness of the data set than conventional geostatistical techniques. The temporal evaluation of the well network is similar in construction to the spatial method, illustrating the analogy between spatial and temporal statistics. This report describes how the results from the spatial and temporal analysis can be used to develop evaluation criteria for a well monitoring network.

Description of Study Area

Broward County is in southern Florida and encompasses an area of about 1,220 mi^2. The study area contains a highly regulated network of manmade canals and is limited to the part of Broward County east of Water-Conservation Areas 2A, 2B, and 3A (fig. 1). The Biscayne aquifer is the principal surficial aquifer in Broward County. This sole-source aquifer (Federal Register Notice, 1979) consists of highly permeable limestone and calcareous sandstone that have hydraulic conductivities of about 1,000 ft/d or more (Fish, 1988). Hydraulic conductivities often exceed 10,000 ft/d in solution-riddled formations, which is typical of southern and eastern Broward County.

The highly regulated network of manmade canals within the study area (fig. 1) supplies water to and drains water from the Biscayne aquifer, thus, producing a major effect on the ground-water flow system. The canal levels are regulated by numerous gates and pumps that provide indirect control of the ground-water levels. Pumping at major municipal well fields also affects the ground-water levels.

The 36 wells in the Broward County ground-water level monitoring network (fig. 2) are equipped with continuous monitors that record water levels at hourly intervals. The well numbers, site identification, period of record, well and casing depths, and casing diameters are listed in table 1. The data collected at these wells, all completed in the Biscayne aquifer, are used to determine daily maximum water levels and are stored by the USGS in a computerized data base. For most of the wells, the period of record was from October 1973 to January 1991, constituting a contemporaneous data base for ground-water levels in the study area.

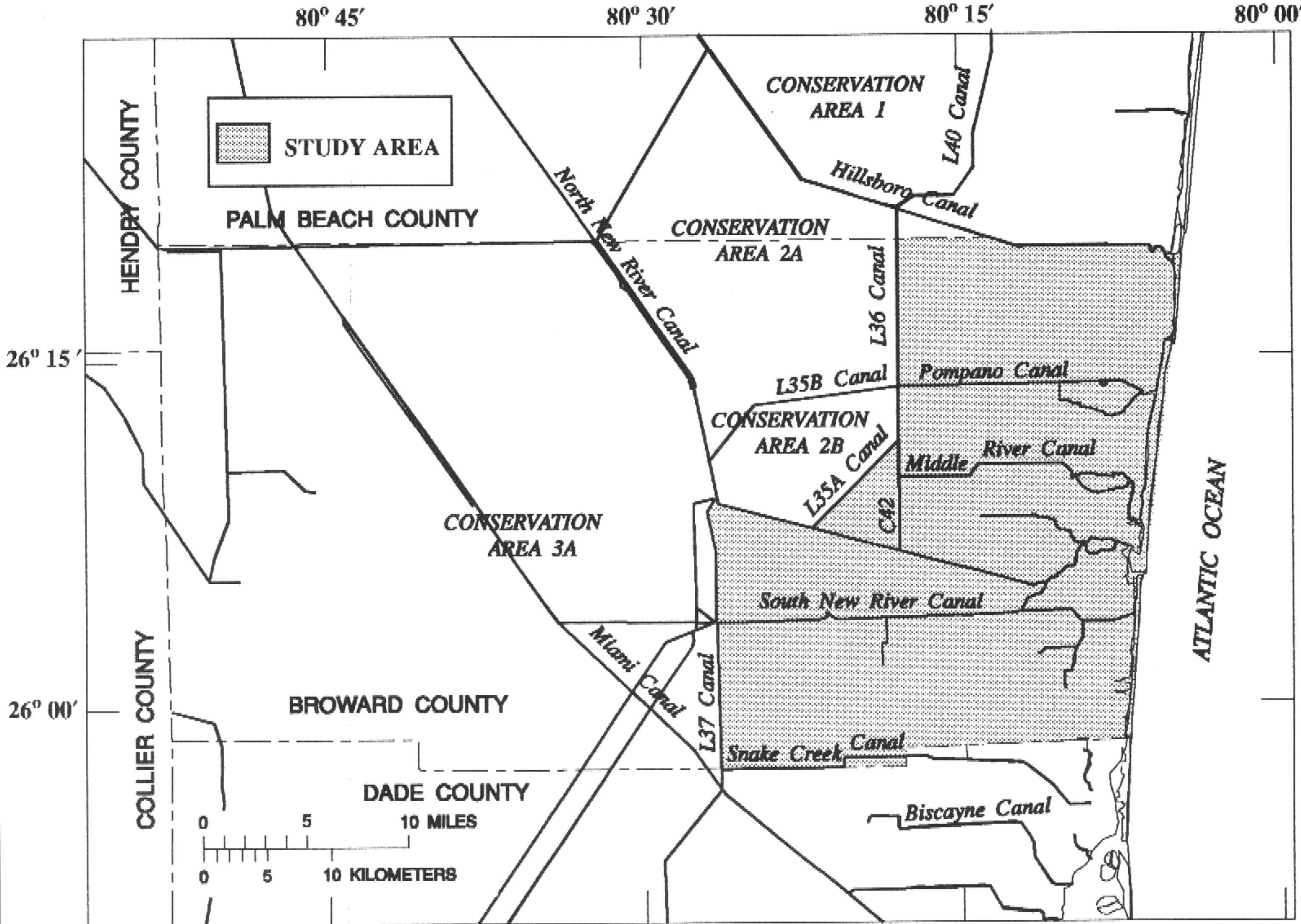
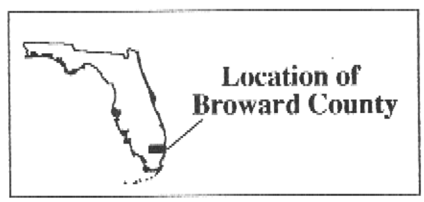


Figure 1. Broward County and the study area.



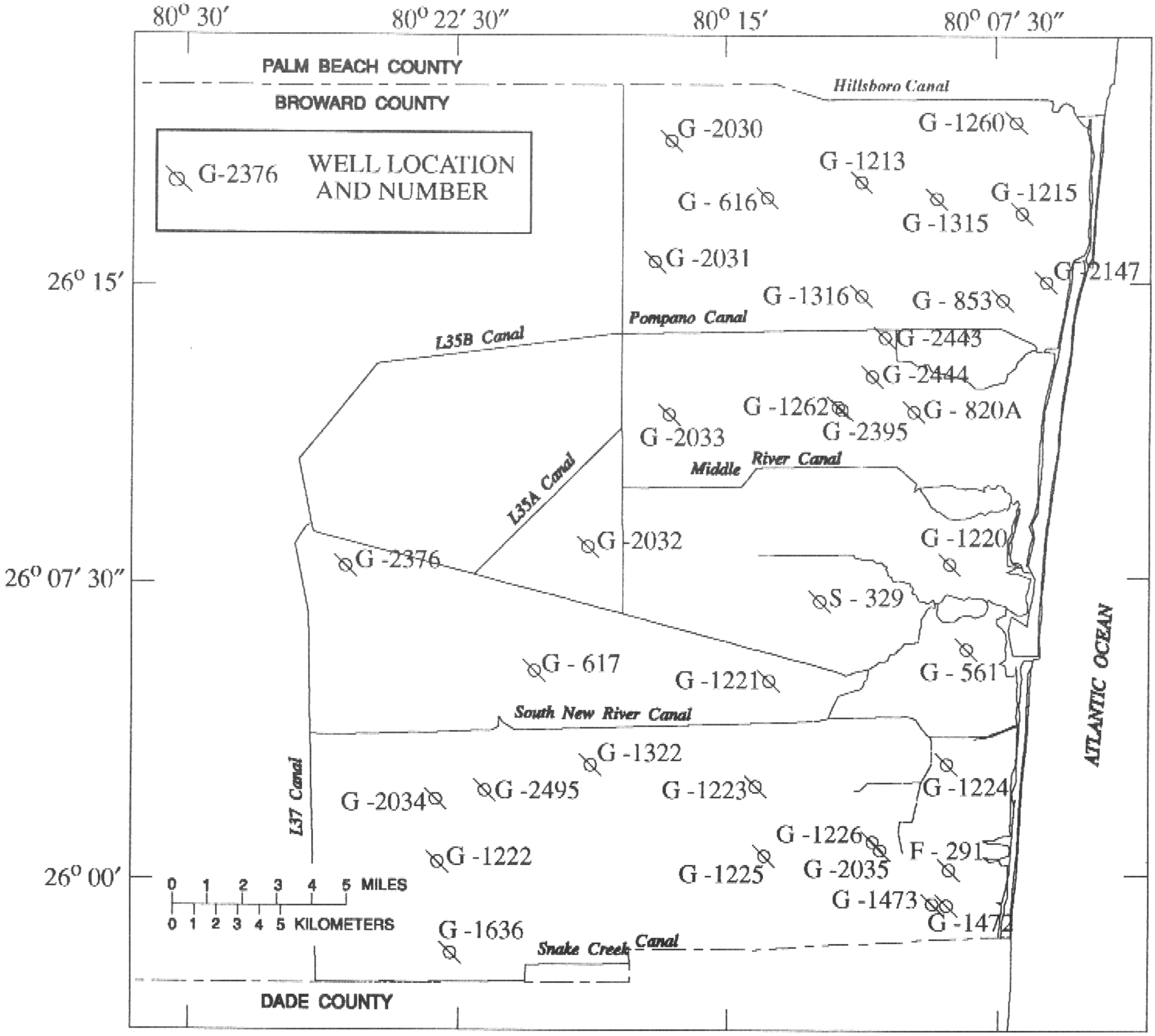


Figure 2. Location of wells in the Broward County ground-water level monitoring network.

Table 1. Well inventory data for the Broward County ground-water level monitoring network

Well number (fig. 2)	Site identification	Period of record	Well depth (feet)	Casing depth (feet)	Casing diameter (inches)
F-291	260010080085001	10-01-73 - 01-01-91	107	107	6
G-561	260545080082001	10-01-73 - 01-01-91	20	20	6
G-616	261710080135001	10-01-73 - 01-01-91	15	15	6
G-617	260515080202101	10-01-73 - 01-01-91	29	28	6
G-820A	261144080094601	10-01-73 - 01-01-91	100	99	4
G-853	261434080071901	10-01-73 - 01-01-91	27	27	4
G-1213	261734080111301	10-01-73 - 01-01-91	20	11.5	5
G-1215	261645080064701	10-01-73 - 01-01-91	20	14	5
G-1220	260752080084701	10-01-73 - 01-01-91	20	20	5
G-1221	260458080134801	10-01-73 - 01-01-91	20	12	5
G-1222	260025080230401	10-01-73 - 01-01-91	20	11.5	5
G-1223	260219080141101	10-01-73 - 01-01-91	20	12	5
G-1224	260252080085301	10-01-73 - 01-01-91	20	12	5
G-1225	260032080135701	10-01-73 - 01-01-91	20	11	5
G-1226	260053080105701	10-01-73 - 01-01-91	20	14	5
G-1260	261903080065601	10-01-73 - 01-01-91	90	90	6
G-1262	261152080115201	10-01-73 - 01-01-91	19	19	7
G-1315	261708080090801	10-01-73 - 01-01-91	14	14	4
G-1316	261441080111301	10-01-73 - 01-01-91	15.5	15.5	4
G-1322	260253080184801	10-01-73 - 10-11-89	13	13	4
G-1472	255916080085401	01-01-74 - 01-01-91	20	20	4
G-1473	255918080091801	10-01-73 - 01-01-91	132	132	8
G-1636	255807080224301	10-01-73 - 01-01-91	24	24	6
G-2030	261837080163001	10-01-73 - 09-30-89	22	21	6
G-2031	261534080165801	10-01-73 - 01-01-92	22	21	6
G-2032	260821080185101	10-01-73 - 01-01-91	22	21	6
G-2033	261141080163401	10-01-73 - 01-01-91	23	21	6
G-2034	260653080184901	10-01-73 - 01-01-91	22	21	6
G-2035	260040080104401	10-01-73 - 01-01-91	52	50	4
G-2147	261501080060701	10-01-74 - 01-01-91	16	16	6
G-2376	260753080253701	06-20-84 - 01-01-91	15	14	4
G-2395	261147080114501	02-09-84 - 01-01-91	80	78	2.5
G-2443	261337080103401	01-01-87 - 01-01-91	145	66	8
G-2444	261238080105501	01-01-87 - 01-01-91	150	83	8
G-2495	260215080214501	02-08-90 - 01-01-92	20	7.1	6
S-329	260657080122301	10-01-73 - 01-01-91	68	68	4

Background

Hydrogeologic, statistical, and linear programming techniques are used in the design of a ground-water network. Loaiciga and others (1992) surveyed approaches to ground-water quality monitoring design; they emphasized the importance of defining the objectives and appropriate methods and surveyed hydrogeologic, statistical, numerical simulation, and variance-based approaches.

One variance-based approach to ground-water quality monitoring network design is the optimization method (Loaiciga and others, 1992). With this method, the network design is posed as a mathematical programming problem, which has an objective function (for example, minimizing the estimation variance) and constraints (governing flow equations, statistical accuracies, and areal coverage of the network). The disadvantages to the optimization method are the simplifications of the hydrogeologic setting that are necessary to convert the parameters to constraints and the lack of consideration for interactions of ground-water monitoring variables, such as surface-water controls and aquifer injection or pumping parameters. These factors are accounted for implicitly in purely statistical analyses.

STATISTICAL METHODS

Analyses using the parameters described in subsequent sections of this report are divided into the spatial and temporal methods. The spatial method provides a graphical depiction of the confidence polygon (CP), a polygon around each well whose boundary is composed of the radial distances to which values can be estimated within a certain accuracy. These values are compared to evaluate the spatial adequacy of the ground-water level network. The temporal method uses the autovariogram to define measurement periods for given accuracies at specific sites.

Network Evaluation

The basic statistical tools used in this ground-water level network evaluation and the measurement scheme design are the variogram and correlation functions. Both functions have their counterparts in the temporal domain (signified by the prefix auto-). These functions represent the degree of similarity between two sets of data or between different parts of the same data set. In the temporal and spatial methods, both functions are based on the variance and covariance of the sets of data and are consequently interrelated.

The basic statistics used to define a single set of data are the mean and the variance. The arithmetic mean is considered the most useful form of the mean in hydrology (Yevjevich, 1982) and is:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

where $\bar{x}$ is the mean, N is the total number of data points, and x_i is the value of the *ith* observation of x. If x is some quantity at a fixed location, x_i is its value.

The variance, σ_x^2, which is always equal to or greater than zero, describes the dispersion of the data and is:

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N-1} (x_i - \bar{x})^2 = \frac{1}{N-1}\sum_{i=1}^{N-1} (X_i^2) - \bar{x}^2 \qquad (2)$$

When two sets of data, x and y, are being compared, a dimensionally similar function defines their linear dependence. The covariance is:

$$\begin{aligned}\sigma_{xy} &= \frac{1}{N-1}\sum_{i=1}^{N-1} (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{N-1}\sum_{i=1}^{N-1} (x_i y_i) - \bar{x}\bar{y}\end{aligned} \qquad (3)$$

where σ_{xy} is the covariance between x and y. If x and y are replaced by points x and $x{+}\tau$ in the same set of data with a fixed time lag τ between them, $\sigma_{x,y+\tau}$ is the autocovariance and is a function of the time lag (τ) between points. Unlike the variance, the covariance and autocovariance can be negative.

Parameters for Defining Spatial Variations

The parameters considered for this report that define spatial variations in a ground-water level network are the correlation coefficient and the variogram. If the covariance is divided by the square root of the product of the variances, a dimensionless parameter results, the correlation coefficient ρ:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}} \tag{4}$$

ρ has a value of 1 for perfectly correlated data sets, 0 for linearly uncorrelated values, and -1 for perfectly correlated values. The standard t-test uses the correlation coefficient and the number of data points to define a parameter t (Yevjevich, 1982):

$$t = \frac{\rho\sqrt{N-2}}{\sqrt{1-\rho^2}} \tag{5}$$

The parameter t follows the Student distribution (Beyer, 1987) with (N - 2) degrees of freedom. The percent area under the curve at a given value in the Student distribution is the confidence level of the significance of the correlations between the two data sets. Although the correlation coefficient is an indication for the linear dependence between data sets and yields a statistic for significance of correlation, it does not quantify the variation between data in a form easily used.

One of the main objectives of a spatial statistical analysis is the definition of the variation between measurements at two points. This can be used to determine the accuracy of values estimated at one measurement point from the data at the other measurement point. The difference between the value in data set x at time i and the value in data set y at time i is referred to as x_i - y_i, which will be designated as δ_i. Thus, the mean value of δ_i is:

$$\bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}\delta_i = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i) = \bar{x} - \bar{y} \tag{6}$$

The variance of the quantity δ_i, as in equation 2, is:

$$\Psi = \frac{1}{N-1}\sum_{i=1}^{N-1}(\delta_i - \bar{\delta})^2 = \frac{1}{N-1}\sum_{i=1}^{N-1}(\delta_i^2) - \bar{\delta}^2 \tag{7}$$

Ψ is referred to as the variogram. With the definitions of δ_i and $\bar{\delta}$ and expanding the quadratic expression, equation 7 becomes:

$$\Psi = \frac{1}{N-1}\sum_{i=1}^{N-1}[(x_i - y_i)^2] - (\bar{x} - \bar{y})^2$$
$$= \frac{1}{N-1}\sum_{i=1}^{N-1}(x_i^2 - 2x_iy_i + y_i^2) - (\bar{x}^2 - 2\bar{x}\bar{y} + \bar{y}^2) \tag{8}$$

With rearrangement of equation 8, the result is:

$$\Psi = \left[\frac{1}{N-1}\sum_{i=1}^{N-1}(x_i^2) - \bar{x}^2\right] + \left[\frac{1}{N-1}\sum_{i=1}^{N-1}(y_i^2) - \bar{y}^2\right] - 2\left[\frac{1}{N-1} - \sum_{i=1}^{N-1}(x_iy_i) - \bar{x}\bar{y}\right] = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} \tag{9}$$

The variogram is the variance of the differences between measurements x and y. The statistical distribution that has a variance equal to Ψ describes the uncertainty in values estimated for one sample based on values only at the other. A common function used in spatial statistics is the semivariogram $\gamma = \Psi/2$ (Cooper and Istok, 1988), which is the arithmetic average variance minus the covariance. The semivariogram is a function of spacing and, theoretically, is zero at zero spacing and approaches the average variance at large spacings. Although the definitions of the variogram and semivariogram vary (Delhomme, 1978; Gambolati and Volpi, 1979; Chirlin and Dagan, 1980; Cooper and Istok, 1988; and Dagan, 1989), the definitions above will be used in this report.

There are several theoretical equations describing the semivariogram as a function of spacing. Because the empirical semivariogram can only be determined at discrete measurement points, an equation must be selected to estimate the value of the semivariogram at intermediate points. The proper form for semivariogram equations based on stochastic flow equations has been researched (Chirlin and Dagan, 1980). Several accepted forms exist (Gambolati and Volpi, 1979). One such equation is the spherical:

$$\gamma = \sigma^2\left[\frac{1.5\xi}{a} - 0.5\left(\frac{\xi}{a}\right)^3\right] + c \qquad for\ \xi < a\ only \tag{10}$$

where ξ is spacing, a is an empirical constant, and c is the variance due to uncorrelated errors called the "nugget." This function is depicted in figure 3. Two more theoretical semivariogram equations are the exponential:

$$\gamma = \sigma^2\left[1 - e^{-\frac{\xi}{a}}\right] + c \tag{11}$$

and the Gaussian (fig. 3):

$$\gamma = \sigma^2\left[1 - e^{-\left(\frac{\xi}{a}\right)^2}\right] + c \tag{12}$$

Theoretically, the value of c is zero, but often uncorrelated errors in field data make a positive c value applicable (Dagan, 1989).

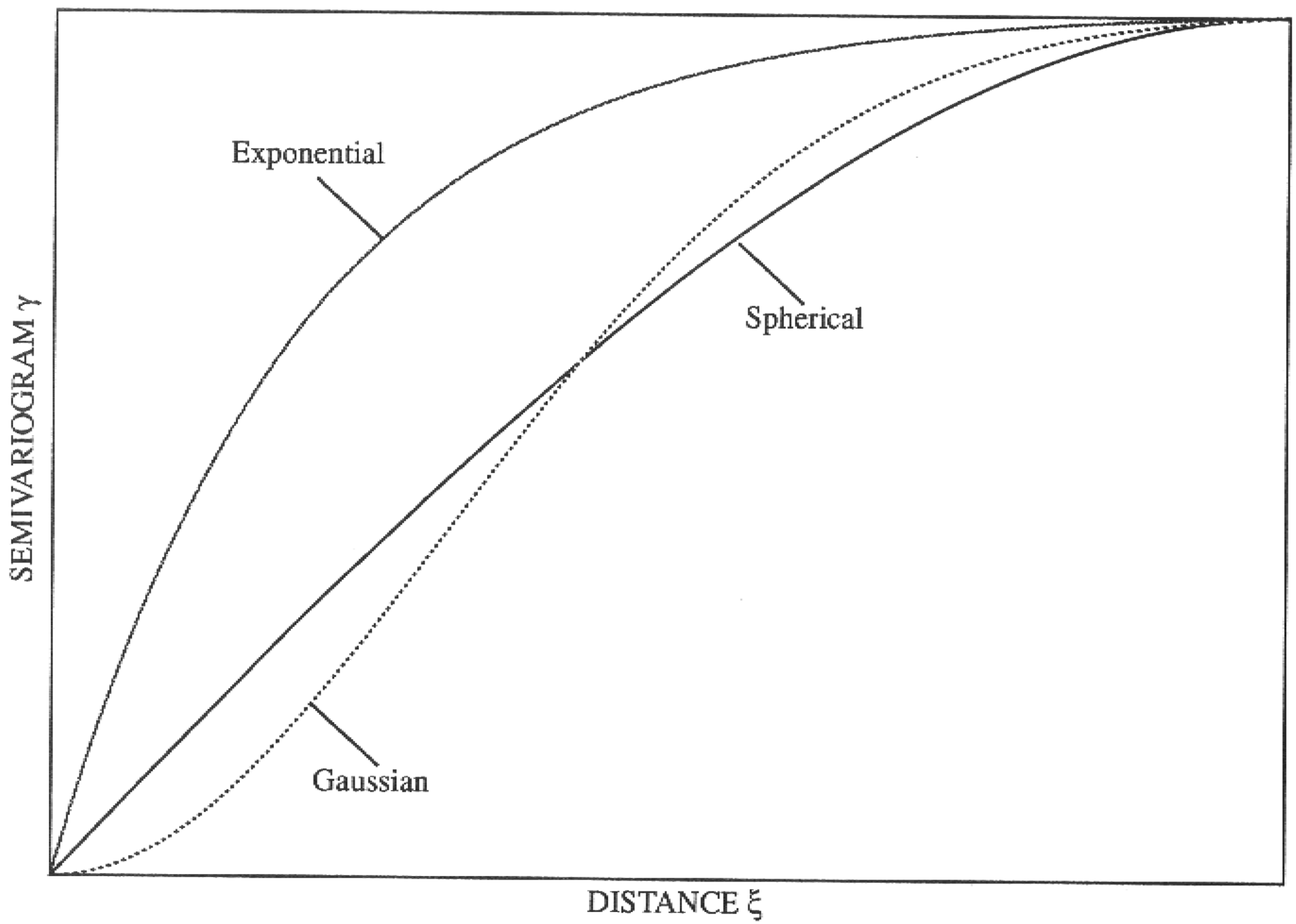


Figure 3. Semivariogram functions.

Parameters for Defining Temporal Variations

The similarities between spatial and temporal statistics become apparent when the covariance and autocovariance are compared. The spacing ξ and the time lag τ are analogous parameters in space and time, respectively. Thus, the variogram Ψ as expressed in equation 9 has an analogy in temporal statistics, logically referred to by the authors as the autovariogram Υ:

$$\Upsilon = 2(\sigma_x^2 - \sigma_{x,x+\tau}) \tag{13}$$

where $\sigma_{x,x+\tau}$ is the autocovariance between data at times x and $x+\tau$. As Ψ is a function of ξ, Υ is a function of τ. The autovariogram and the variogram allow determination of the accuracies of estimates.

The autocorrelation coefficient ρ_τ is defined as:

$$\rho_\tau = \frac{\sigma_{x,x+\tau}}{\sigma_x^2} \tag{14}$$

The relation between the variance, autocovariance, autocorrelation coefficient, and autovariogram is identical to the relation between the variance, covariance, correlation coefficient, and variogram.

The parameters ρ, Ψ, and γ yield information as to the correlation between values at wells and the accuracies of estimates between wells. The parameters ρ_τ and Υ yield information about the correlation between values at an individual well taken at different times and the accuracies of estimates at one time based on known values at another time. These statistical parameters are used to determine redundancy of wells, locations where wells are needed, and optimum measurement intervals at each well.

Spatial Method

A method commonly used in ground-water network analysis for water quality and water levels is Kriging (Olea, 1975). This method produces a statistical uncertainty in estimates of unmeasured sites as functions of distance between measurement location and determining a best estimate at unmeasured locations by the averages of the values at known points weighted inversely to their uncertainty at the unknown point by minimizing the Kriging variance. Kriging can be applied to network design by finding the point with maximum uncertainty based on the assumed statistical uncertainty functions, the semivariograms (Cooper and Istok, 1988).

Kriging has been incorporated into larger schemes for network design. A random search method can be used to create a series of alternative network designs. The design with the least uncertainty in estimates is considered optimal (Journel and Huijbregts, 1978). Rather than design a new network, an existing network can be improved by adding measurement locations at the points of minimum calculated estimation error to reduce the uncertainty by a desired amount (Delhomme, 1978; McBratney and others, 1981; and de Marsily, 1986). Kriging can be combined with mathematical programming by using the Kriging uncertainty as the objective function and the Kriging system equations as the constraints (Bardossy and Bogardi, 1983; Carrera and others, 1984).

Details that are not measured with the applications of Kriging for network design are implicitly estimated, and no additional information can be derived than inherent in the original data (Delhomme, 1979). Thus, the uncertainty at a proposed measurement point is estimated based on the values at known points and assumed statistical functions (the variogram) between the points. This can result in significant errors for sparse networks (Seo and others, 1990). When adding proposed estimation locations to an existing network, the accuracy of the entire network is assumed to improve. However, real data at these proposed estimation locations do not exist, and the improvement in the network must also be an estimate.

Errors of estimation are inevitable in Kriging because of the limited information in the field data. An approach that defines sparse networks using the regional effectiveness of each measurement location would more reasonably reflect the lack of data. Also, the approach should reflect that the network is too sparse to make interpolated estimates as in Kriging, and an estimate is only made by measuring a nearby well. Kriging defines each needed measurement location as a point. Besides incorrectly inferring that the location can be accurately defined on such a precise scale, even in a sparse network, standard Kriging does not include all factors, such as physical obstructions, legal permissions, and local phenomena, which determine an actual measurement location. These considerations provide a new statistical scheme for yielding general information for a sparse network.

In a sparse network, Kriging can be considered inappropriate because of insufficient data. An appropriate scheme would delineate the area around each well within which water levels can reasonably be predicted by the well. Thus, the user admits the network is too sparse to interpolate between wells as in Kriging. Water levels near each well location can be estimated with greater accuracy than water levels at points farther from the well because nearby values tend to be more alike than distant values. This property can be statistically defined using information from all other wells so radial distances can be defined from each well to points of given estimation confidence. The boundary connecting the endpoints of these radial lines encloses an area representative of the spatial monitoring confidence of the well. Such delineation of a CP for each measurement location would be made according to the criteria of prediction within certain limits at a given level of confidence. Instead of defining the accuracy required as a Kriged error (or uncertainty variance), it can be stipulated as specific error limits on the measurement itself within a level of confidence. This is a quantity more amenable to measurement criteria.

Because the boundary of the CP follows the individual radial distances determined by the statistical relations between the water levels at the location of interest and all other measurement locations, a regional form of Kriging is not used and smaller scale variations are retained. Locations where additional measurements are needed are indicated by areas between the CPs. Areas where CPs overlap indicate effective redundancy in measurements between wells. Varying the accuracy criteria varies the size of the CPs; CP sizes would also vary with the season considered in the analysis. The higher recharge variability of the wet season causes more water-level fluctuations than in the dry season. Additionally, municipal well pumpage varies with the season with a subsequent effect on ground-water level fluctuation.

The basis of the spatial analysis is to calculate the semivariogram for each pair of wells. Unlike the technique used in conventional Kriging, the semivariogram is not evaluated as a single function of distance and angle but is retained as a separate function for each pair of wells. Thus, the individual idiosyncracies that relate to each pair of wells are not lost in a generalized variogram function for the area. This is an attempt to make the most of a sparse network. This analysis reduces the generalization in the method and allows the detection of local discontinuities. A Geographic Information System (GIS) interface allows graphical representations of the results for evaluation.

The data for each pair of wells are initially sorted by time to ensure that a one-to-one correspondence exists. For each comparison, only the days that have data at both wells are used. The software package STATIT[1] (Statware Inc., 1990) contains specific routines to calculate statistical parameters. Programs are used to call these STATIT routines when needed. These programs are listed in appendix 1. The mean, variance, correlation coefficient, and variogram are calculated for each well pair. The standard t-test (eq. 5) is performed on the correlations for each pair. If there is no significant correlation between the wells, the variogram is not applicable.

If a statistical distribution can be fit to the differences in water levels at each well, confidence intervals can be predicted at any location between the wells and extrapolated beyond the wells. A distance from each well can be calculated that has the value of the semivariogram γ_d which indicates a desired confidence in estimates made at the point. This method for a hypothetical well network and a CP for one well (no. 1) in the network is shown in figure 4.

To determine the semivariogram value for the desired confidence, the statistical distribution of the differences in well water levels must be defined. Most statistical quantities in hydrology can be defined by the normal or log-normal distributions. The normal distribution as applied to this situation is:

$$f(\delta_i) = \frac{1}{\sqrt{\Psi 2\pi}} e^{-\frac{(\delta_i - \bar{\delta})^2}{2\Psi}} \qquad (15)$$

where $f(\delta_i)$ is the density function value at $\delta = \delta_i$. This is the normal distribution of the differences between x and y, which has a mean of $\bar{\delta}$ and a variance Ψ. The natural log-normal distribution that applies to this situation is:

$$f(\delta_i) = \frac{1}{\delta_i \sqrt{\Psi_L 2\pi}} e^{-\frac{(ln\delta_i - \bar{\delta}_L)^2}{2\Psi_L}} \qquad (16)$$

where $\bar{\delta}_L$ is the mean of the values of $ln(\delta_i)$ and Ψ_L is the variance of the values of $ln(\delta_i)$.

---

[1]The use of firm, trade, manufacturer, and product names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

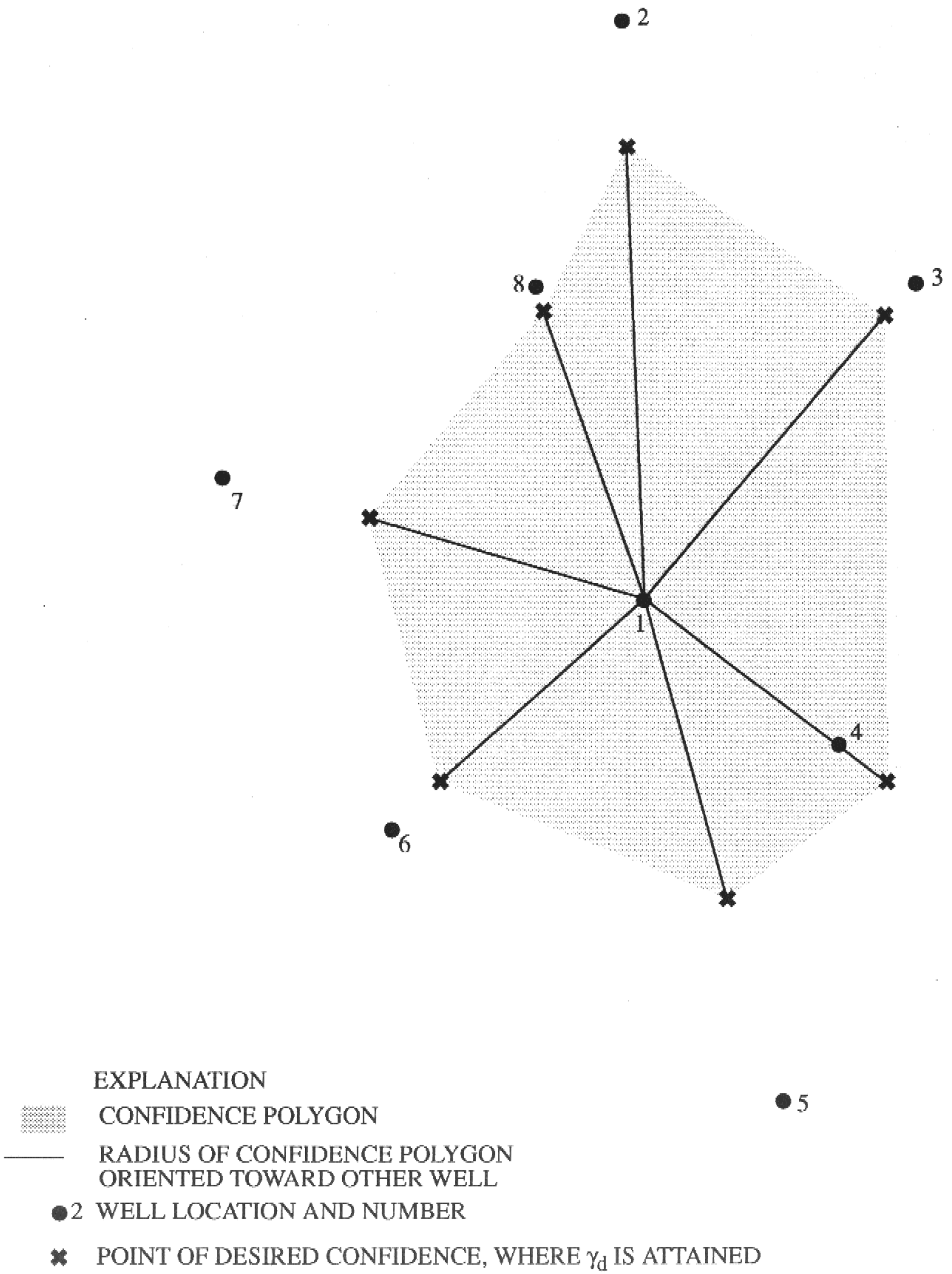


Figure 4. Hypothetical well network with polygon connecting points of equal desired confidence.

To determine which distribution (eq. 15 or 16) should be used, the Shapiro-Wilk test for normality (Shapiro and Wilk, 1965) is performed on the values of δ_i and of $ln(\delta_i)$. This test is implemented using a STATIT routine. The test indicates which distribution (eq. 15 or 16) best represents the data. After the determination has been made, a maximum allowable error in estimation ε = desired accuracy in predicting water level in well y from well x must be selected. Additionally, a confidence level P = percent of data that can be predicted at well y by well x within maximum allowable error ε must be selected. This maximum allowable error ε is equal to the maximum allowable absolute value of $\delta_i - \bar{\delta}$ for the accuracy of prediction desired. For the selected probability distribution, a value of Ψ is determined so that the percentage of the distribution curve area between ordinate values of $+\varepsilon$ and $-\varepsilon$ is equal to P. This is determined using the standard normal variate Z which, in the case of the normal distribution, is:

$$Z = \frac{(\delta_i - \bar{\delta})_{max}}{\sqrt{\Psi}} = \frac{\varepsilon}{\sqrt{\Psi}} \qquad (17)$$

and for the log-normal distribution, the derivation in appendix 2 yields:

$$Z = \frac{(ln\delta_i - \bar{\delta}_L)_{max}}{\sqrt{\Psi_L}} = \sqrt{ln\left(\frac{\Psi}{\bar{\delta}^2}+1\right)}\left[\frac{ln\left(\frac{\varepsilon+\bar{\delta}}{\bar{\delta}}\right)}{ln\left(\frac{\Psi}{\bar{\delta}^2}+1\right)}+\frac{1}{2}\right] \qquad (18)$$

Standard statistic tables (Snedecor and Cochran, 1982) give values under the normal curve for values of Z. Thus, for the desired value of P (area under normal curve) and for an ε, a value of Z yields the necessary value of Ψ. This value, divided by two, yields the semivariogram value γ_d.

The value of the variogram between each pair of wells is calculated according to equation 9. Dividing this value by two gives the value of the semivariogram corresponding to the distance between the two wells γ_f. To define the semivariogram as a function of distance for all intermediate points between the wells, one of the empirical models as described in equations 10, 11, and 12 must be selected. To select the model, the values of the semivariogram as a function of distance to each well are plotted to determine the general shape of the curve.

After equations 10, 11, or 12 are selected as the "best" semivariogram representation, the equation is applied along each of the radial lines to other wells (fig. 4). Values of c and a are selected in the equation so that the semivariogram attains the value γ_f determined from the field values at the distance of the other well. Because the coefficient c is ideally zero, this leaves the calculation of the value of a only. A different value of a is calculated for each radial line from a well. The semivariogram equation can be used to determine the value of distance $\xi=h$. This is the distance to the point having the desired value of the semivariogram γ_d.

A separate value of h is determined for each radial line (fig. 4). The points on each radial line, a distance (h) away from well 1, are connected to form a boundary that encloses the CP as shown in figure 4. For example, there are eight wells in figure 4, so there are seven well pairs associated with well 1 (well 1 with wells 7 and 8), and there will be seven values of γ_f calculated. This is done after calculating statistics and applying the t-test for correlations on each pair for many days. Several limitations of this polygon depiction are:

1. It is assumed that the radial distance to the edge of the CP changes linearly between known points on the edge of the CP. If a large angle exists between two radial lines, there is substantial error of estimation in the intervening region.
2. A distant well, such as well 2 (fig. 4), might correlate with the central well, but a hydrologic condition in the intervening area, such as a well field, might indicate that ground-water levels in the polygon cannot be accurately predicted. This is an inherent problem with conventional Kriging as well as any statistical analysis. Variogram relations with closer wells could detect this problem by defining the variograms to these closer wells.
3. Because of the highly correlated wells, the semivariogram equation could indicate that the distance to the boundary of the CP is farther away than the well itself. This situation is depicted between wells 1 and 4 in figure 4. The semivariogram relation between wells 1 and 4 does not indicate what occurs beyond well 4, so it can be erroneous to extend the polygon beyond this well.
4. There is no proof that the "best" semivariogram function applies between wells where no data exist. The function can only be selected on the basis of data at the existing well distances.
5. The wells at the outer edges of the network have no data for statistical relations exterior to the network for determination of CP radii, so the CPs of these wells tend to be

truncated on their sides facing outward from the network. However, if the exterior wells are considered the boundary of the network, the CP area outside the network does not contribute to the network.

6. Whichever statistical distribution is selected as the best representation of the data, the data cannot be expected to fit it without some error.

Despite these limitations, there are several advantages to the polygon depiction that make it very useful for assessing the adequacy of a sparse network. These advantages are:

1. The polygon gives a good visual description of the best well to predict water levels in an approximated area (assuming no interpolation between wells) and the exterior areas where predictions cannot be made.
2. Redundant wells can be identified by overlapping polygons.
3. Areas that need wells to accurately define the water levels can be easily located by gaps between the polygons.
4. When a well has a missing record and data from nearby wells are needed to estimate this record, the polygons indicate which adjacent well is best suited for this purpose. For example, in figure 4, well 4 best approximates the values for well 1.
5. Wet and dry seasons can be analyzed separately because wells redundant in one season might not be redundant in the other season.
6. By retaining individual semivariogram functions along each radial line for each well, the polygon method should be more sensitive to spatial anisotropy. In Kriging, assuming a general anisotropy would be difficult in a sparse network.

This process has been applied to define CPs for the wells in the study area. Distances (h) from a well in opposite directions along the same line are the same length because they use the same semivariogram function. The analysis is made for all data in the specified analysis period and for the wet and dry seasons.

Pairs of wells with each CP overlapping the other well location can be considered "effectively redundant." This indicates that the water levels in each well can be used to predict the water level at the location of the other well within the specified criteria. This does not necessarily indicate that removal of either one of a pair of redundant wells will yield the same amount of field information. Even when wells have CPs that overlap well locations, there would remain some areas covered by only one CP for the well.

Temporal Method

In regional time-varying statistics, an analogy exists between the analytical methods used in the spatial and temporal domains. When concerned with the spacing between measurement points, the parameters used in Kriging define the corresponding spatial variations. Instead of defining the spatial variations, the temporal variations are defined, and these parameters can be used to define the required time intervals between measurements. The time period between measurements at one location can be the temporal equivalent to the spacing between measurement points.

The temporal analysis is simpler than the spatial analysis in several aspects. Only operating in one dimension (time), anisotropy is not a factor as it is in the spatial domain with two or three dimensions. Thus, a single temporal function can be defined for each measurement point. Also, by making measurements at regular time intervals, developing a uniform data set for analysis is less costly and less labor intensive in most field conditions than setting up measurement locations in a uniform spatial grid.

The effect of temporal measurement intervals is as important as spatial measurement location optimization. Even with a spatial measurement network that is determined to predict with a certain accuracy, no accuracy can be attained if the temporal measurement interval is too long. A given spatial network can only be used to predict accurately with a determined temporal measurement interval. The spatial and temporal accuracies are very dependent on each other.

The optimum measurement periods for each well are defined in terms of an autovariogram function. The autovariogram for each well is calculated according to equation 13 for varying lag times τ. Because the STATIT package calculates the autocovariance $\sigma_{x,x+\tau}$ and autocorrelation ρ_τ, these functions are used to determine the autovariogram by:

$$\Upsilon = 2\sigma_{x,x+\tau}\left(\frac{1}{\rho_\tau} - 1\right) \tag{19}$$

Combining equations 13 and 14 yields equation 19. A plot of the autovariogram Υ against time lag τ yields the value of time lag beyond which the autovariogram rises above the specified criteria. This time lag corresponds to the optimum measurement period.

The specified criteria for the minimum allowable autovariogram value should relate to the accuracy.

The variogram values for the spatial analysis can be used from equation 17 or 18 as the autovariogram criteria. The same accuracy for spatial prediction is also applicable when determining measurement periods.

For the period of record of each well, the autovariogram as a function of time lag is determined by equation 19. Because the field data are at 1-day intervals, the time lag must be defined in 1-day increments. Values for the autovariogram are linearly interpolated between daily values. The value of time lag when the autovariogram first exceeds the criteria is noted as the maximum interval between measurements. The analysis is made for the same uncertainty criteria used in the spatial analysis.

Evaluation Criteria

The spatial and temporal analyses yield information which can be applied to a ranking system evaluation criterion to determine the relative importance of each well in the ground-water level network. A type of ranking system, DRASTIC (Aller and others, 1985), has been used to evaluate the ground-water pollution potential of a region. The approach used in DRASTIC to develop a ranking system has been applied to the problem of ground-water level network design. DRASTIC considers these factors: D (depth to water), R (net recharge), A (aquifer media), S (soil media), T (topography), I (impact of the vadose zone), and C (conductivity of the aquifer). Each factor is rated from 1 to 10, which is then weighted according to importance. The sum of the weighted ratings is the DRASTIC parameter. This method can be used to determine the pollution potential of any hydrogeologic setting. Details of the rating method are described by Aller and others (1985). The technique has been applied to the southern Florida area (Herr, 1990). A similar method using more complex alphanumeric codes has been used to evaluate the potential of ground-water contamination from waste disposal sites (LeGrand, 1983).

A similar technique was developed in this study for ranking the importance of each well in a ground-water level network. Called Spatial and Temporal Adequacy and Redundancy Evaluation (STARE), the relevant data are CP area, effective redundancy, and temporal confidence intervals. The first two parameters result from the spatial analysis, and the third parameter results from the temporal analysis. The CP area depicts the effective area monitored by the well, the effective redundancy depicts the tendency to monitor an area covered by another well, and the temporal confidence interval depicts the necessary frequency of measurements at the well or the rate of the temporal fluctuation at the well.

The STARE approach normalizes the three parameters and calculates a weighted sum to depict the relative importance of each well to the network. In the DRASTIC scheme, some of the parameters are nonnumeric (soil type, aquifer type, and impact of vadose zone), so a 1 to 10 scale rating for each parameter is used. However, in the STARE scheme, each parameter is numeric. Normalizing the parameters is a simple and logical approach. The determination of what is "good" or "bad" in each parameter is another important consideration which must be defined so that the weighted sum of the parameters is a logical representation.

The normalizing and weighting schemes that have been selected for STARE are not the only variables that affect the final results. The criteria selected when determining the CPs, redundancies, and temporal confidence intervals (with allowable errors ε and percent confidence P) greatly affect the results. If the criteria were made strict enough, there would probably be no redundant wells, and the redundancy would have no effect on the STARE results. The most important criteria indicate some redundancies in the network but do not involve a large number of the wells. This should correspond to some significant accuracy desired in the network. The selection of the criteria has a degree of inherent subjectivity.

The first parameter, CP area, is logically normalized in relation to the total area covered by the network. Thus, each CP is expressed as a fraction of the network area that each well covers: a number between 0 and 1. The proportionality between the CPs of different wells should not drastically change when the accuracy criteria or the season considered for the analysis are changed, so data from one analysis are considered representative.

The second parameter is the redundancy. The redundancy value is the number of times a well is part of a redundant pair is dimensionless, so it can be considered to be normalized against a single redundancy. It is logical to include several analyses with differing accuracy criteria and season considered in the development of this parameter because wells might be required to have differing accuracy or be used for seasonal analyses. The redundancy is a sum of several analyses and would increase nonlinearly relative to well proximity (a well redundant in one analysis is likely to be redundant in another analysis). Thus, it severely penalizes the most redundant wells. The redundancy value can vary from 0 to (number of analyses) × (number of wells - 1).

A more serious consideration is the simplest one; that is, determining which well is more valuable--a well with a large CP or a well with a small CP. The large CP seems to be the obvious choice, retaining the wells that represent the most area. However, the wells with smaller CPs should be considered because of the uniqueness of the data collected there. A small CP indicates that data at a well cannot be predicted by data collected elsewhere. A rationale to resolve this conflict is to consider that the second parameter, redundancy, is a representation of the tendency to mimic other wells. A well with a large CP that is not redundant is obviously valuable. If the CP is large and the well is highly redundant, the redundancy parameter can serve to penalize the well. If the CP is small and redundant, the well should be penalized more than a well with a large CP and the same redundancy. Therefore, it is logical to give more value to larger CPs and use the redundancy as a counterweight.

The third parameter, temporal confidence interval, is the most difficult to normalize. The only representative timescale for the system is the existing measurement interval (in this example, 1 day). When dividing by this number, values tend to be greater than one with no theoretical upper limit (other than the period of record). This must be considered when weighting this parameter.

The weighting scheme for the three normalized parameters in the STARE scheme is based on three factors: (1) the adequacy of the spatial coverage of the well network, (2) the redundancy of wells in the network, and (3) optimal well measurement intervals for the wells in the network. The three normalized STARE parameters (CP area, redundancy, and temporal confidence interval) address each factor, respectively. Thus, an equal weighting is indicated. The normalized CP area is a fraction less than or equal to one, whereas the redundancy and temporal confidence interval are often greater than one. This justifies expressing the CP area as a percentage (multiplying by 100).

The normalized and weighted scheme resulting from these criteria is:

$$STARE\ coefficient = \left(\frac{confidence\ polygon\ area}{total\ area\ covered\ by\ network} \times 100\right) - redundancy - \frac{temporal\ confidence\ interval}{sampling\ interval} \quad (20)$$

For a set of wells in a network, the lowest STARE coefficient can be subtracted from all the STARE values for convenience. This adjusted STARE coefficient has a lowest value of zero and has the same proportionality as the unadjusted STARE coefficient.

If a well is identified as the least important to the network (lowest STARE coefficient) and is removed, the analysis must be repeated for the remaining wells to obtain a new STARE coefficient ranking for the network. The individual CP areas, total area covered by the network, and the redundancy change, but the temporal confidence intervals for the remaining wells are the same.

SPATIAL ANALYSIS

Basic water-level statistics for the wells in the study area indicate the size and variability of the data. These statistics are given in table 2. The mean levels and variances greatly vary between wells, which is mainly due to regional effects of well fields, canals, and coastal water levels. Appendix 2 lists the FORTRAN programs and Unix scripts used to apply the following procedure. The correlation coefficients were determined and variogram analyses were performed for four periods, using the:

- Period of record (October 1, 1973-January 1, 1991) with $\varepsilon = 0.5$ ft, $P = 90$ percent,
- Period of record (October 1, 1973-January 1, 1991) with $\varepsilon = 0.3$ ft, $P = 90$ percent,
- Wet seasons only (May 1-October 31) with $\varepsilon = 0.3$ ft, $P = 90$ percent, and
- Dry seasons only (November 1-April 30) with $\varepsilon = 0.3$ ft, $P = 90$ percent.

The results were used to determine the statistical distribution with which to represent the data and the representative semivariogram function.

Table 2. Water levels and statistics for wells in the Broward County ground-water level monitoring network

Well No.	Days of data	Water levels, in feet above and below (-) sea level			Variance (feet squared)	Standard deviation (feet)
		Minimum	Mean	Maximum		
F-291	6,302	0.24	1.56	5.52	0.44	0.66
G-561	6,302	.39	1.62	4.52	.38	.62
G-616	5,461	5.38	8.61	13.41	2.06	1.43
G-617	6,302	2.51	3.80	6.85	.33	.57
G-820A	4,046	-3.04	3.16	7.69	3.71	1.93
G-853	6,247	-6.62	.06	5.51	3.54	1.88
G-1213	6,302	9.39	12.31	14.99	1.04	1.02
G-1215	6,302	-2.20	2.67	8.50	4.84	2.20
G-1220	6,302	.43	1.81	6.37	.45	.67
G-1221	4,878	.75	1.94	6.52	.32	.56
G-1222	6,122	1.59	3.70	6.59	.59	.77
G-1223	6,302	1.43	2.46	6.39	.33	.57
G-1224	6,301	-.22	1.64	5.14	.48	.69
G-1225	6,135	.91	2.40	6.54	.60	.77
G-1226	6,302	.30	1.53	8.97	.43	.65
G-1260	6,302	-.71	2.88	7.59	2.57	1.60
G-1262	5,771	-11.19	-5.37	3.05	10.16	3.19
G-1315	6,302	6.26	10.00	13.66	1.55	1.25
G-1316	4,701	6.69	8.52	12.78	.77	.88
G-1322	5,855	2.05	3.26	5.84	.21	.46
G-1472	6,210	.23	1.60	5.33	.53	.73
G-1473	6,302	.27	1.50	5.37	.42	.65
G-1636	6,302	1.79	2.99	5.01	.27	.52
G-2030	5,780	6.09	8.19	10.99	.69	.83
G-2031	6,667	4.75	7.29	10.96	.30	.55
G-2032	6,302	2.85	4.36	7.25	.46	.68
G-2033	6,190	5.21	6.60	10.01	.31	.56
G-2034	6,302	1.49	3.66	6.19	.50	.71
G-2035	6,301	.22	1.51	5.52	.38	.61
G-2147	5,937	-1.02	1.90	6.53	1.63	1.28
G-2376	2,387	3.95	6.09	7.23	.47	.69
G-2395	2,292	-13.76	-7.21	-.82	5.80	2.41
G-2443	1,462	4.45	5.94	8.35	.45	.67
G-2444	1,462	.46	2.33	5.02	1.01	1.01
G-2495	667	2.48	3.65	5.32	.39	.62
S-329	6,295	-1.26	1.27	5.13	1.08	1.04

Period of Record Analysis

The number of daily data points for concurrent days between wells using the period of record, October 1, 1973 to January 1, 1991 (not counting those with no concurrent data), ranges from 266 days for well pair G-2495/G-1262 to 6,302 days for wells F-291, G-561, G-617, G-1213, G-1215, G-1220, G-1223, G-1226, G-260, G-1315, G-1473, G-1636, G-2031, G-2032, and G-2034. Well G-2495 has no concurrent data with wells G-1322 and G-2030 (table 1). Therefore, correlations or variograms could not be calculated for well pairs G-2495/G-1322 and G-2495/G-2030.

The correlation coefficients were determined for well pairs containing concurrent data. Values ranged from 0.019 for well pair G-561/G1262 to 0.980 for well pair F-291/G-1473. The t-test according to equation 5 was performed, and the two-tailed probability (probability that the correlation coefficient is not significant) was obtained from the Student distribution. Well pairs G-561/G-1262 and G-2032/G-1262 failed to meet the 90 percent confidence limits with a two-tailed probability of 0.147 and 0.109, respectively, and were not used in the variogram analysis. The majority of well pairs had a two-tailed probability of 0.000, indicating that the correlations were highly significant.

The Shapiro-Wilk test indicated little difference in fitting the differences in residuals to a normal or a log-normal distribution. The best fit to a normal distribution and a log-normal distribution had an R-squared value of 0.998 for well pair G-2032/S-329. The worse fits were 0.589 for normal data and 0.591 for log-normal data for well pair G-1226/G-2035. The normal and log-normal R-squared values had the same mean, within 0.961. The normal distribution was selected to use for predictions because of the close results between the normal and log-normal distributions and the lack of sensitivity of the scheme to the exact distribution.

To obtain an initial estimate for the best semivariogram function to use in this area, the semivariograms as a function of distance were calculated for each well related to all other wells. The resulting scatter plots are shown in figure 5. These plots cannot be used to "best fit" semivariogram functions for two reasons. First, the values of the semivariograms in the plots are much higher than those of concern, so any fit to these values would not be reliable at the small semivariogram values of interest. Second, the semivariogram values on any one plot occur for wells at differing directions, so fitting to the points would ignore anisotropy. The plots can be used as a rough guide to the general shape of the semivariogram functions. None of the plots seem to have the flat low end of the curve with increasing slope that the Gaussian function has in figure 3. Wells G-616, G-853, G-1215, G-1260, and G-2395 indicate the steep low-end slope of the spherical and exponential equations in figure 3. The spherical and exponential equations seem to better express the variogram functions in this area than the Gaussian equation. However, all three equations were used for comparison.

The semivariogram values γ_f were calculated for each well pair from the variance and covariance data. Based on the semivariogram achieving this value at the distance between the wells, values for the coefficient a in equations 10, 11, and 12 were calculated. A different value of a is used then for each radial line to another well.

The desired accuracy in predicted head difference ε must be selected as a realistic value. The predicted head difference does not have to be as accurate as the well measurements because ε indicates a confidence in estimation for a region. Ground-water level contour maps of Broward County (Lietz, 1992) were drawn with 1.0-ft contours for most of the county. To predict values from such a map, an accuracy of $\varepsilon = 0.5$ ft with a reasonable strict confidence limit was needed. Standard statistics indicate that for a 90 percent confidence limit for normally distributed data, the value of the standard normal variate $Z = 1.6448$. Using these values ε of Z in equation 17, it is indicated that head prediction within 0.5 ft with 90 percent confidence can occur at a distance in which the variogram Ψ is 0.0924 ft^2 or the semivariogram γ_d is 0.0462 ft^2.

The radial distance along each line to the boundary of the CP is calculated according to the spherical, exponential, and Gaussian semivariogram equations. Expressing the radial distances as a percent of the total distance between each well pair, the average differences in radial distances showed the exponential equation produced results that were 2.4 percent lower than the spherical equation case, and the Gaussian case produced results that were 14.2 percent higher than the spherical equation. The close results of the exponential and spherical equations are significant, considering that examination of figure 5 and comparison with the functions in figure 3 indicate that the exponential and spherical equations are better representations of this system than the Gaussian equation. Between the

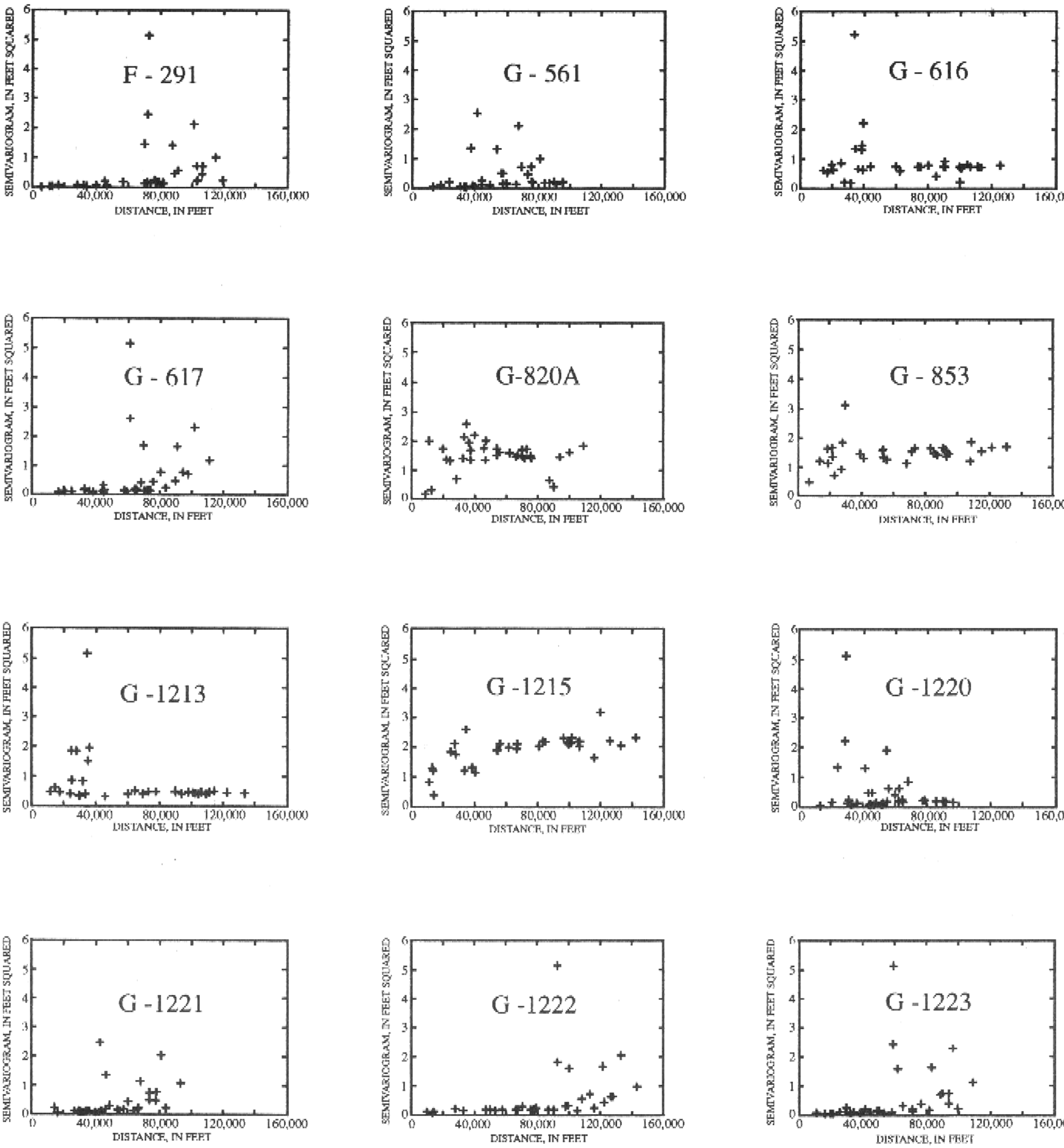


Figure 5. Calculated semivariograms for wells in the Broward County ground-water level monitoring network. Well locations shown in figure 2.

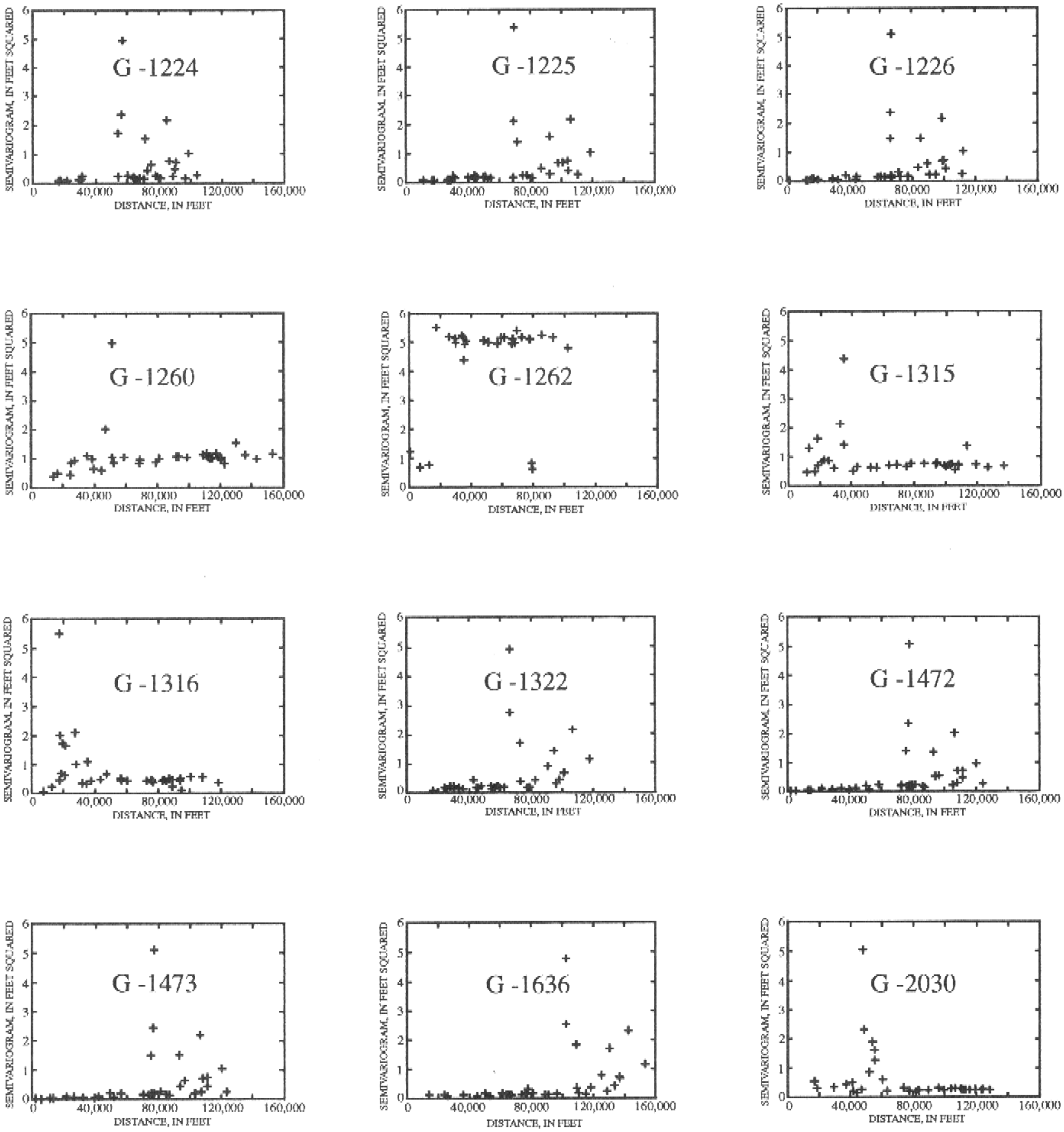


Figure 5. Calculated semivariograms for wells in the Broward County ground-water level monitoring network--Continued. Well locations shown in figure 2.

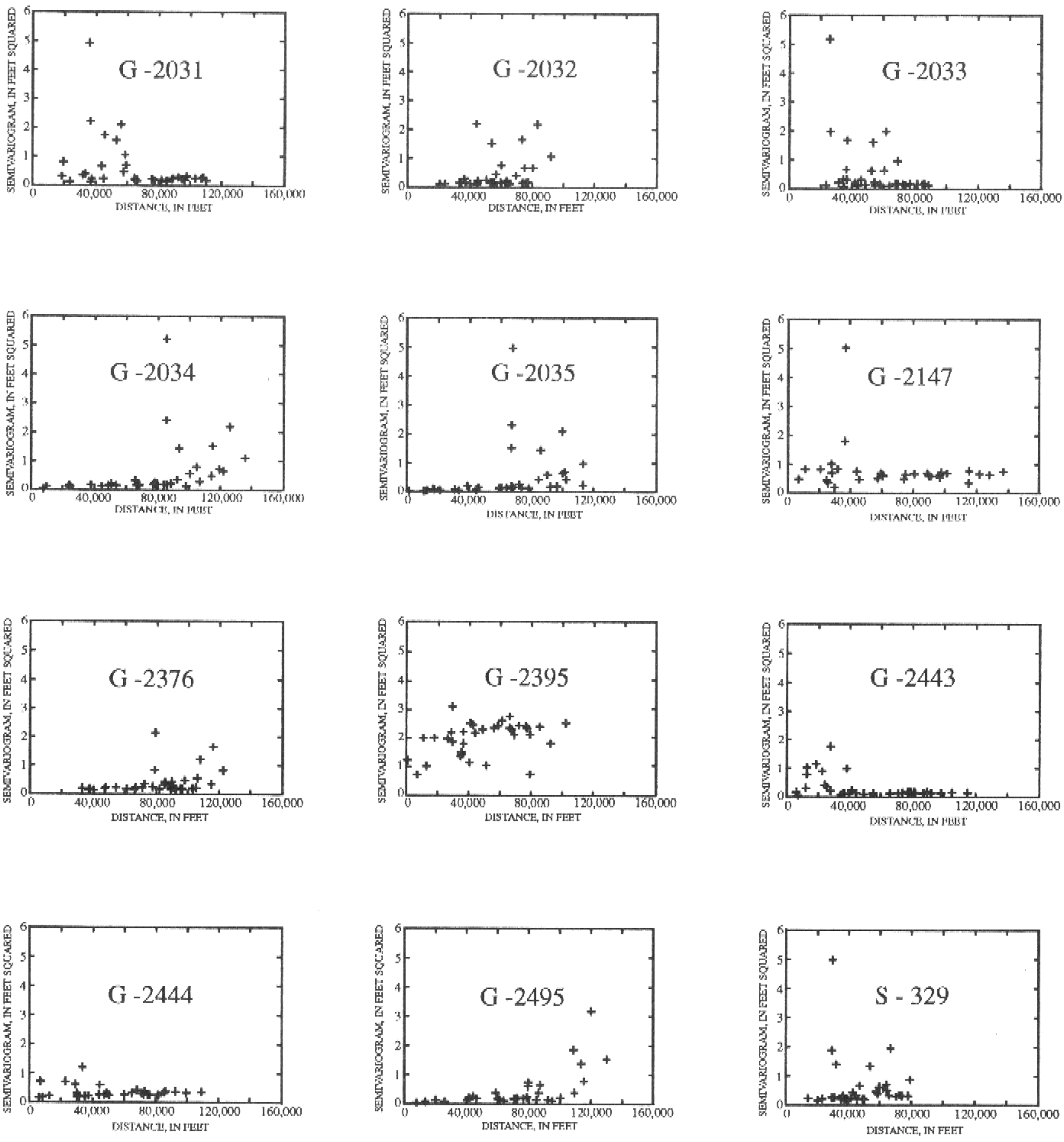


Figure 5. Calculated semivariograms for wells in the Broward County ground-water level monitoring network--Continued. Well locations shown in figure 2.

exponential and spherical equations, the spherical equation was selected to represent the extent of the CPs because its results are between those of the exponential and Gaussian equations, yet close to the exponential equation.

The radial distances were calculated to the points where γ_d is 0.0462 ft^2 using the spherical semi-variogram equation. The resulting CPs are shown in figure 6. The CPs indicate the areas in which each respective well can be used to predict the ground-water level within 0.5 ft at 90 percent confidence. Using equation 17 with the standard normal variate (Z) tables is equivalent to predicting levels within 0.3 ft at 67.6 percent confidence or within 0.1 ft at 25.8 percent confidence.

The individual CPs in figure 6 can be discerned more easily in northern Broward County than in southern Broward. In southeastern Broward County, the CPs are so large and clustered that it is difficult to identify the CP for a particular well. (The CP of a particular well should be displayed alone.) The large CPs indicate high correlation with other wells, which is due to the relatively flat, steady water table in southeastern Broward County. The analysis indicates that water levels can be predicted within the criteria specified for the existing wells virtually everywhere in this area. However, there are two anomalies: the CP of well F-291 that points to the southwest (fig. 6, "a") and the CP of well G-1473 that points to the northeast (fig. 6, "b"). These anomalies result from the high correlation between wells F-291 and G-1473, and not much confidence can be placed in the vicinity of these two wells (see limitation no. 3 in the *"Spatial Method"* section).

The wells with the smallest CPs are in northeastern Broward County (fig. 6). This area is subject to steeper ground-water level gradients (Lietz, 1992). The analysis indicates that each of these wells can only be used to estimate water levels in a small area in close vicinity to the well. Many more wells would be needed in this area to obtain representative regional coverage.

Ground-water levels in southeastern Broward County are well monitored, but there is missing areal coverage between the CPs of wells in other parts of the county. There are large gaps between the CPs of wells G-2032, G-2033, G-1262, G-2395, S-329, and G-1221 in central Broward County and between the CPs of well G-2376 and adjacent wells in west-central Broward County (fig. 6).

The sharp projections on the CPs (fig. 6) result from high correlation between relatively distant wells (see limitation no. 2 in the *"Spatial Method"* section). The appearance of the projection as a spike indicates that adjacent wells have a much lower radius of the CP.

The removal of distant correlations creates a more realistic representation of the CPs as shown in figure 7. The individual CPs are somewhat discernible, but the southeastern coast is still dense. Although this method smooths the CP boundaries, the selection of which projections should be removed is arbitrary. The predicted areal coverage of the CPs does not change appreciably.

To illustrate the effect of making the criteria more strict, the variogram analysis was implemented with a criterion of water-level prediction within 0.3 ft at 90 percent confidence ($\gamma_d = 0.0166$ ft^2). This can be viewed as either tightening the water-level requirement by 0.2 ft or tightening the confidence limit by 9.4 percent (determined by Z and normal table). The results are shown in figure 8. A significant reduction in size of the CPs is observed (figs. 6 and 8), which is expected because the proportionality between the CP radius and water-level criteria is not linear (eqs. 10 and 16). With this criterion of 0.3 ft, relatively few of the CPs overlap, and each well monitors a small area. The individual CPs are clearer in figure 8 than in figures 6 and 7. The CP of a well is generally the same shape when the criteria are changed, only smaller.

Wet-Season Analysis

The correlation coefficients and variogram analyses were conducted using the data for the wet season, defined as the period from May through October. The number of concurrent days for the wet-season analysis ranged from 154 days for well pair G-2395/G-2495 to 3,159 days for wells F-291, G-561, G-617, G-1213, G-1215, G-1220, G-1223, G-1224, G-1226, G-1260, G-1315, G-1473, G-1636, G-2031, G-2032, G-2034, and S-329. For this analysis (as was the case for the period of record analysis), well pairs G-1322/G-2495 and G-2030/G-2495 had no concurrent data and could not be used for the correlation or variogram analyses.

The correlation coefficients for the wet season ranged from 0.006 at well pair G-1262/G-2032 to 0.980 at well pair F-291/G-1473. Nine well pairs showed no significant correlation within a confidence limit of 90

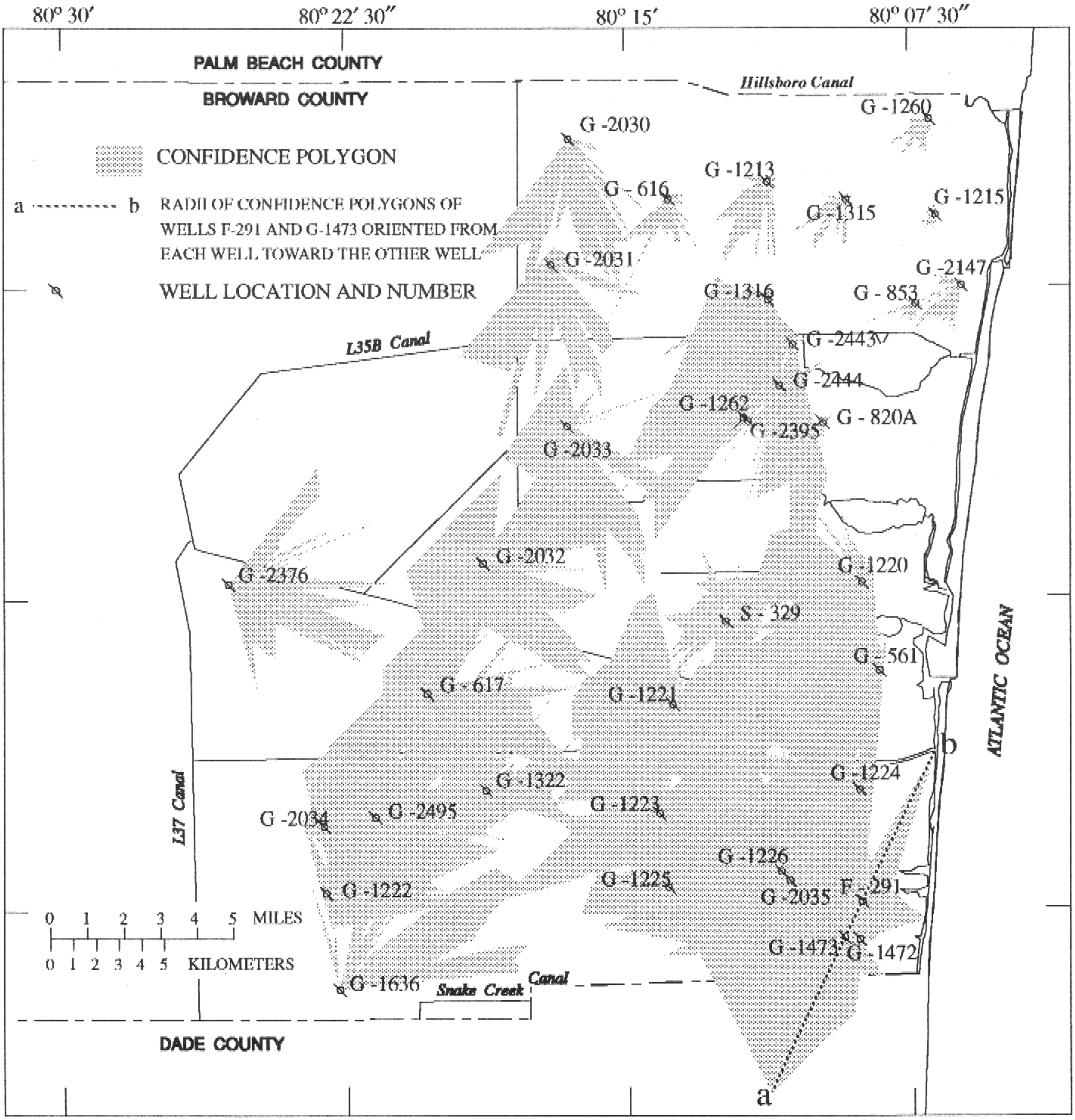


Figure 6. Confidence polygons for 0.5-foot criterion at 90 percent confidence limit.

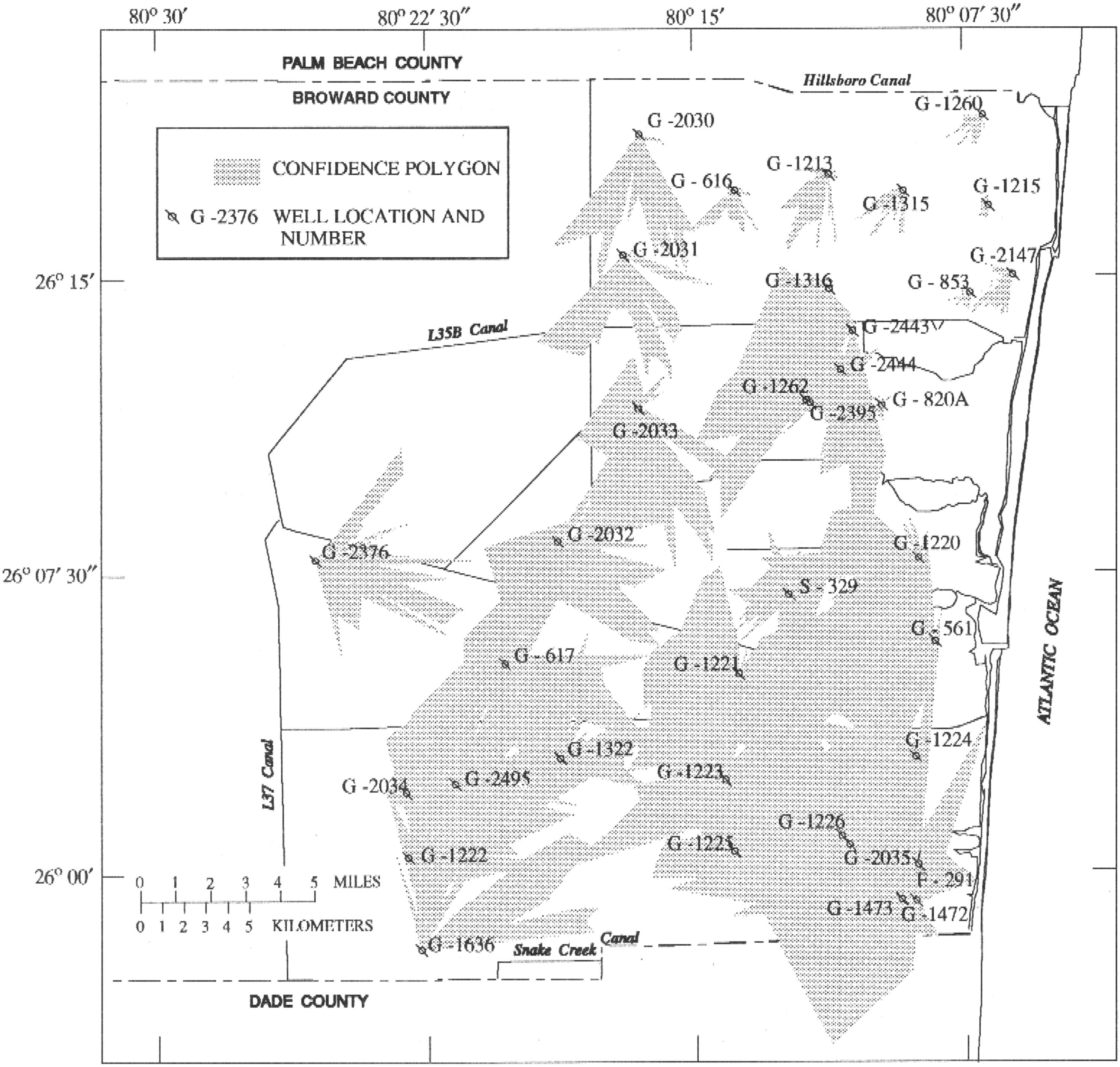


Figure 7. Confidence polygons for 0.5-foot criterion at 90 percent confidence limit with distant correlations removed.

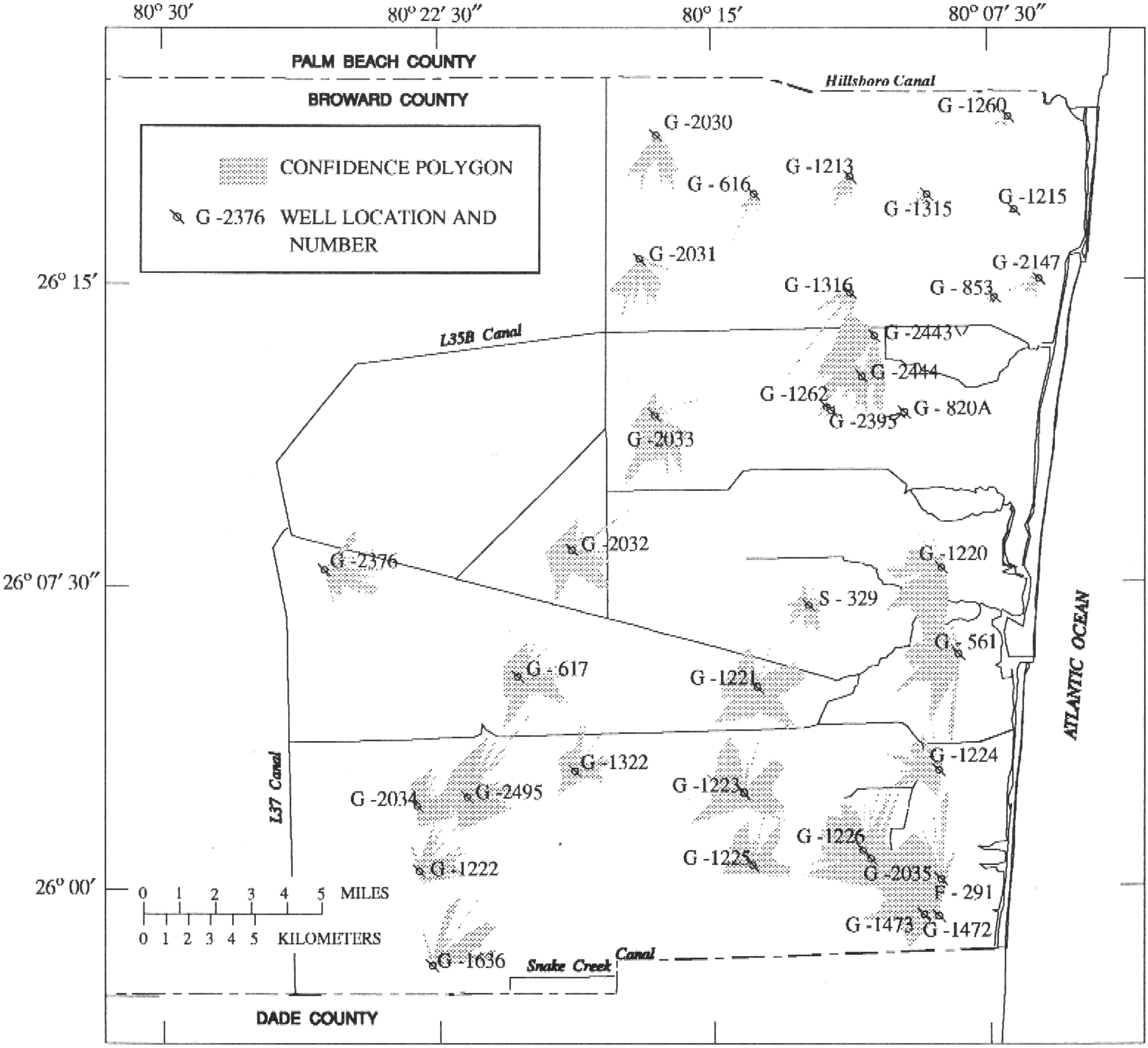


Figure 8. Confidence polygons for 0.3-foot criterion at 90 percent confidence limit.

percent. The two-tailed probabilities for these well pairs are: 0.205 at G-820A/G-1636, 0.103 at G-853/G-1262, 0.356 at G-1215/G-1262, 0.155 at G-1225/G-1262, 0.171 at G-1262/G-2031, 0.749 at G-1262/G-2032, 0.739 at G-1262/G-2033, 0.607 at G-1315/G-1322 and 0.532 at G-1316/G-2147. These well pairs were not used for the variogram analysis.

The CPs for the wet-season analysis using the 0.3-ft criterion at 90 percent confidence are shown in figure 9. Most of the CPs for the wet season are smaller than those for the period of record (figs. 8 and 9). This result is expected, considering that during the wet season water-level fluctuations are smaller due to rainfall recharge.

Dry-Season Analysis

The correlation coefficients and variogram analyses were performed using the data for the dry season, defined as the period from November through April. The number of concurrent days for the dry-season analysis ranged from 82 days for well pair G-1262/G-2495 to 3,143 days for wells F-291, G-561, G-617, G-853, G-1213, G-1215, G-1220, G-1223, G-1224, G-1226, G-1260, G-1315, G-1473, G-1636, G-2031, G-2032, G-2034, G-2035 and S-329. Well pairs G-1322/G-2495 and G-2030/G-2495 had no concurrent data for the dry seasons and could not be used for the correlation or variogram analyses.

The correlation coefficients for the dry season ranged from 0.008 at well pair G-561/G-1262 to 0.971 at well pair F-291/G-1473. There were 10 well pairs with no significant correlation with a confidence limit of 90 percent. The two-tailed probabilities for these well pairs are: 0.664 at G-561/G-1262, 0.156 at G-617/G-1262, 0.452 at G-1221/G-2495, 0.182 at G-1222/G-1316, 0.304 at G-1223/G-2495, 0.315 at G-1262/G-1472, 0.199 at G-1262/G-2147, 0.718 at G-1636/G-2495, 0.917 at G-2376/G-2495, and 0.277 at G-2395/G-2495. These well pairs were not used for the variogram analysis.

The CPs for the dry-season analysis using the 0.3-ft criterion at 90 percent confidence are shown in figure 10. CPs are larger for the dry season than for the period of record (figs. 8 and 10), which corresponds to the smaller spatial variability of this lower recharge period. The long projections in the CP boundaries of wells G-1316 and G-2495 oriented toward each other are pronounced in the dry season.

Well Utility and Redundancy

One method of ranking the utility of each well is by the area of each CP. This is based on the concept that wells with a larger CP can be used to monitor a larger area. The wells are ranked in order of decreasing CP area for the 0.5-ft criterion at 90 percent confidence in table 3, but site-specific interests are not considered in this ranking. However, a well that only monitors a small CP area does so because its fluctuations are higher. Therefore, the wells with the smaller CPs can be considered as important because they are monitoring a highly variable situation. The debate is addressed in the "*Evaluation Criteria*" section.

The effectively redundant well pairs at 90 percent confidence for the 0.5-ft and 0.3-ft analyses are determined by the overlapping of polygons with well locations. Figure 11 illustrates the difference when tightening the water-level fluctuation criteria. Using the 0.5-ft criterion at 90 percent confidence, there are 12 redundant well pairs, including F-291/G-561, F-291/G-1226, F-291/G-1472, F-291/G-1473, F-291/G-2035, G-561/G-1220, G-1226/G-1473, G-1226/G-2035, G-1472/G-1473, G-1472/G-2035, G-1473/G-2035, and G-2034/G-2495. The southeastern corner of the study area contains the greatest well redundancy (fig. 11).

Analysis of redundant pairs for the 0.3-ft criterion at 90 percent confidence for the entire year, wet season only, and dry season only data demonstrates the seasonal difference in redundancy (fig. 12). Well pairs F-291/G-1472 and F-291/G-1473 are redundant with the 0.3-ft criterion at 90 percent confidence as shown in figures 11 and 12. Although well F-291 is effectively redundant with wells G-1472 and G-1473, these two wells are not effectively redundant with each other. This is probably because the ground-water level contours in this area are roughly oriented north to south (Lietz, 1992). Thus, well pairs F-291/G-1472 and F-291/G-1473 are closer to being on the same contour line than well pair G-1472/G-1473.

Only well pair F-291/G-1473 is effectively redundant in the wet-season analysis (fig. 12). The redundancy is much lower than in other situations due to the high variability of the water levels in the wet season.

The eight well pairs redundant in the dry-season analysis only include: F-291/G-1226, F-291/G-1472, F-291/G-1473, F-291/G-2035, G-617/G-2495, G-1226/G-1473, G-1226/G-2035, and G-2034/G-2495. This is

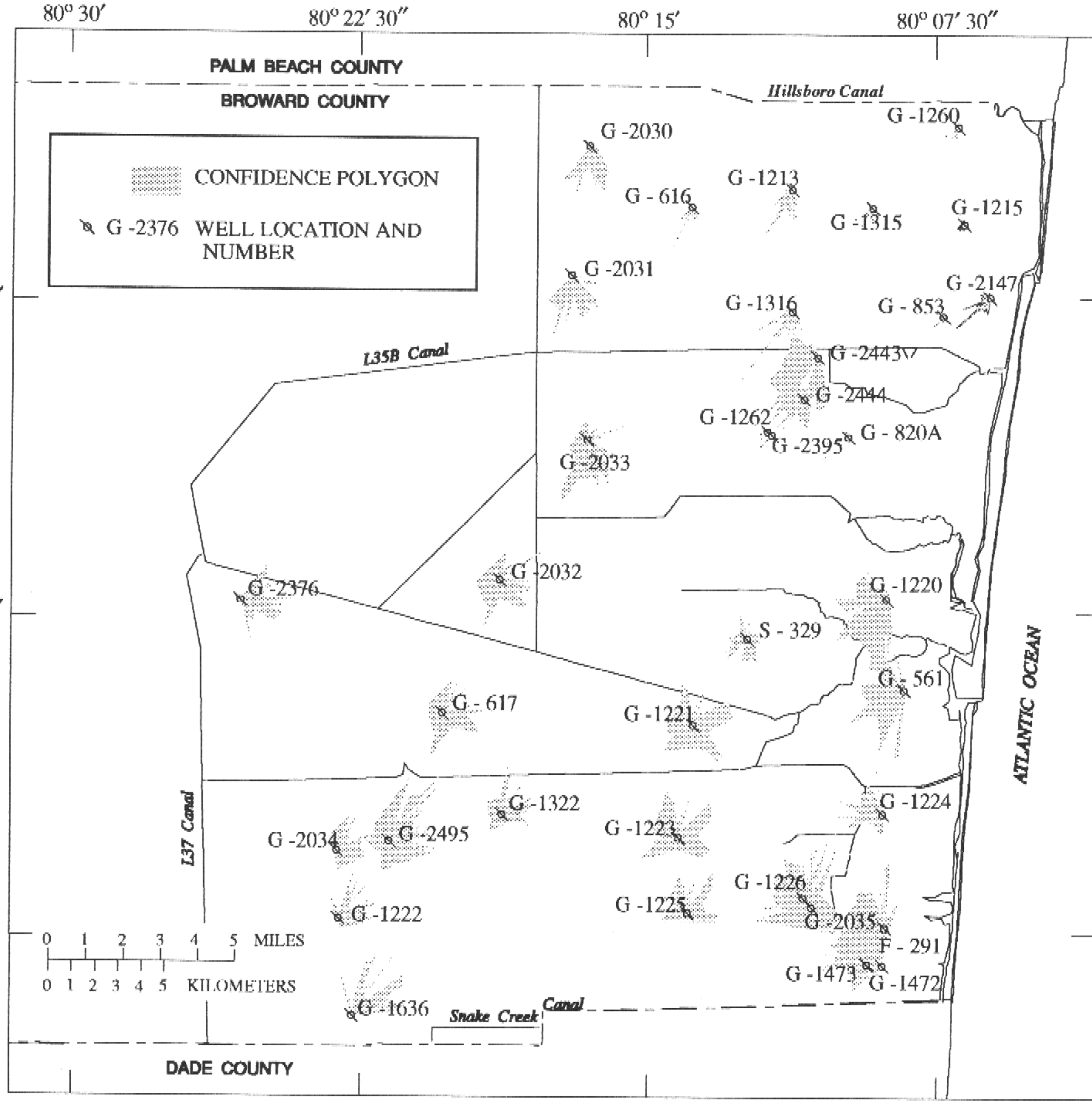


Figure 9. Confidence polygons for 0.3-foot criterion at 90 percent confidence limit using wet-season data only.

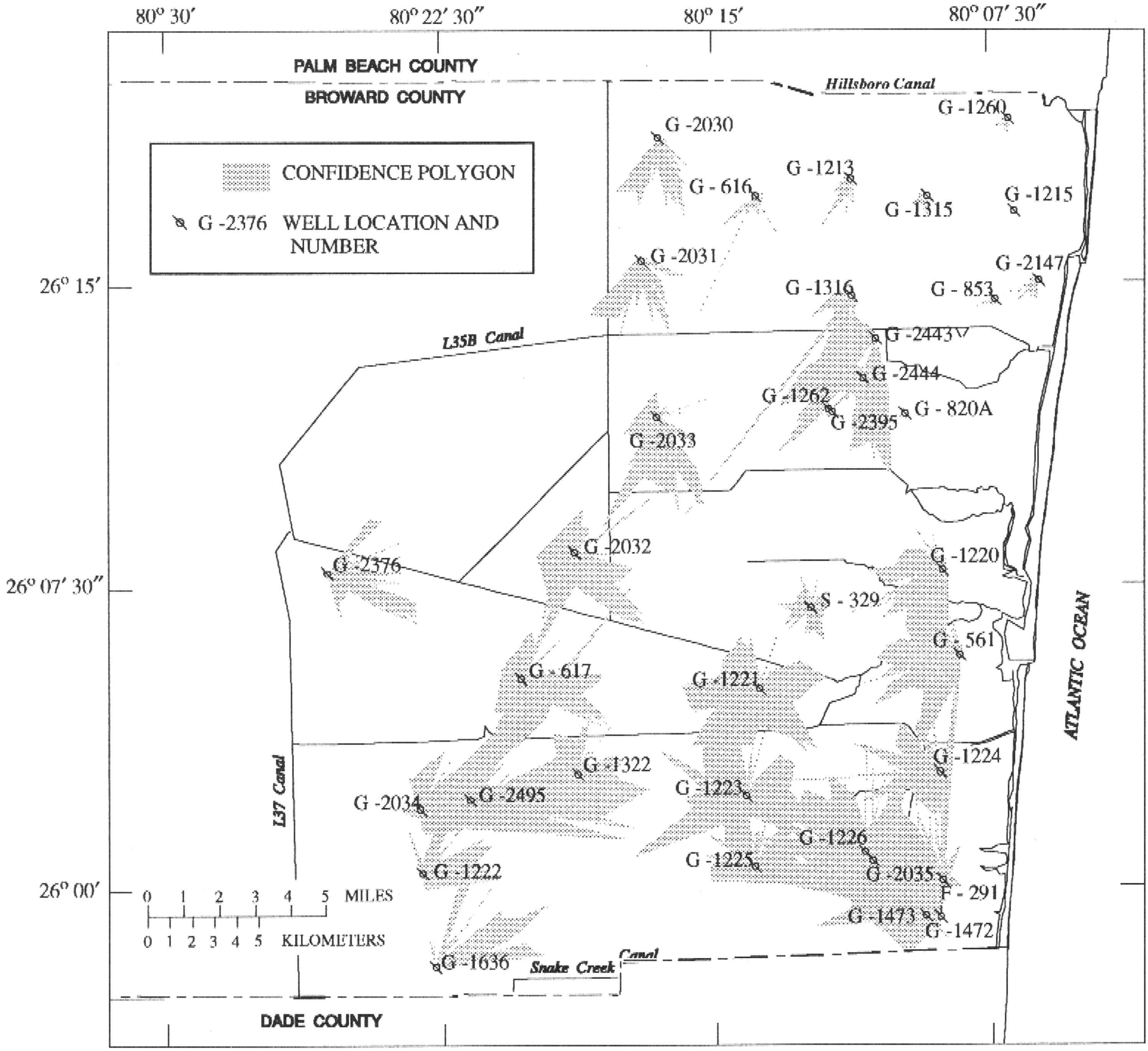


Figure 10. Confidence polygons for 0.3-foot criterion at 90 percent confidence limit using dry-season data only.

Table 3. Wells ranked in order of decreasing confidence polygon area using all data, 0.5-foot criterion at 90 percent confidence limit

Well number	Confidence polygon area (square miles)	Well number	Confidence polygon area (square miles)
F-291	27.17	G-1322	7.74
G-2035	25.08	G-2376	7.69
G-1223	21.56	G-1472	7.34
G-1221	21.42	G-1222	6.75
G-1220	20.50	G-2030	6.19
G-2443	18.73	G-2444	4.49
G-2495	18.65	S-329	4.17
G-561	16.78	G-1316	2.27
G-1226	16.53	G-1213	1.59
G-1473	14.50	G-616	.95
G-2032	13.88	G-2147	.66
G-2033	13.54	G-1315	.64
G-617	13.15	G-1260	.36
G-1225	11.74	G-820A	.22
G-1636	9.83	G-853	.15
G-2034	8.28	G-1215	.11
G-1224	8.11	G-2395	.08
G-2031	7.78	G-1262	.03

the largest number of redundant wells for the 0.3-ft criterion at 90 percent confidence (fig. 12). The lower variability of the water levels during the dry season is the cause of this large number of redundant wells.

Using the 0.3-ft criterion at 90 percent confidence, the following analysis is indicated. Well pair F-291/G-1473 duplicates each other within the criteria at all times. Well pair F-291/G-1472 duplicates within the criteria for yearly analyses but does not duplicate within the criteria for studies of the wet season only. Well pairs F-291/G-1226, F-291/G-2035, G-617/G-2495, G-1226/G-1473, G-1226/G-2035, and G-2034/G-2495 can be considered to represent each other within criteria for the dry season only.

The sum of the redundancy data for the four analyses (0.5-ft criterion, 0.3-ft criterion, wet season, and dry season) indicates that the wells can be ranked in terms of the number of times each well is indicated as part of a redundant pair (table 4). The wells in the southeastern part of the study area have the greatest redundancy. This ranking does not give consideration to site-specific interests, total CP areas, or existing gaps in the network.

Table 4. Ranking of well redundancy using four analyses (0.3-foot criterion, 0.5-foot criterion, wet season, and dry season)

Well number	Number of times well is part of a redundant pair
F-291	12
G-1473	8
G-2035	6
G-1226	6
G-1472	5
G-2495	3
G-561	2
G-2034	2
G-617	1
G-1220	1
All other wells	0

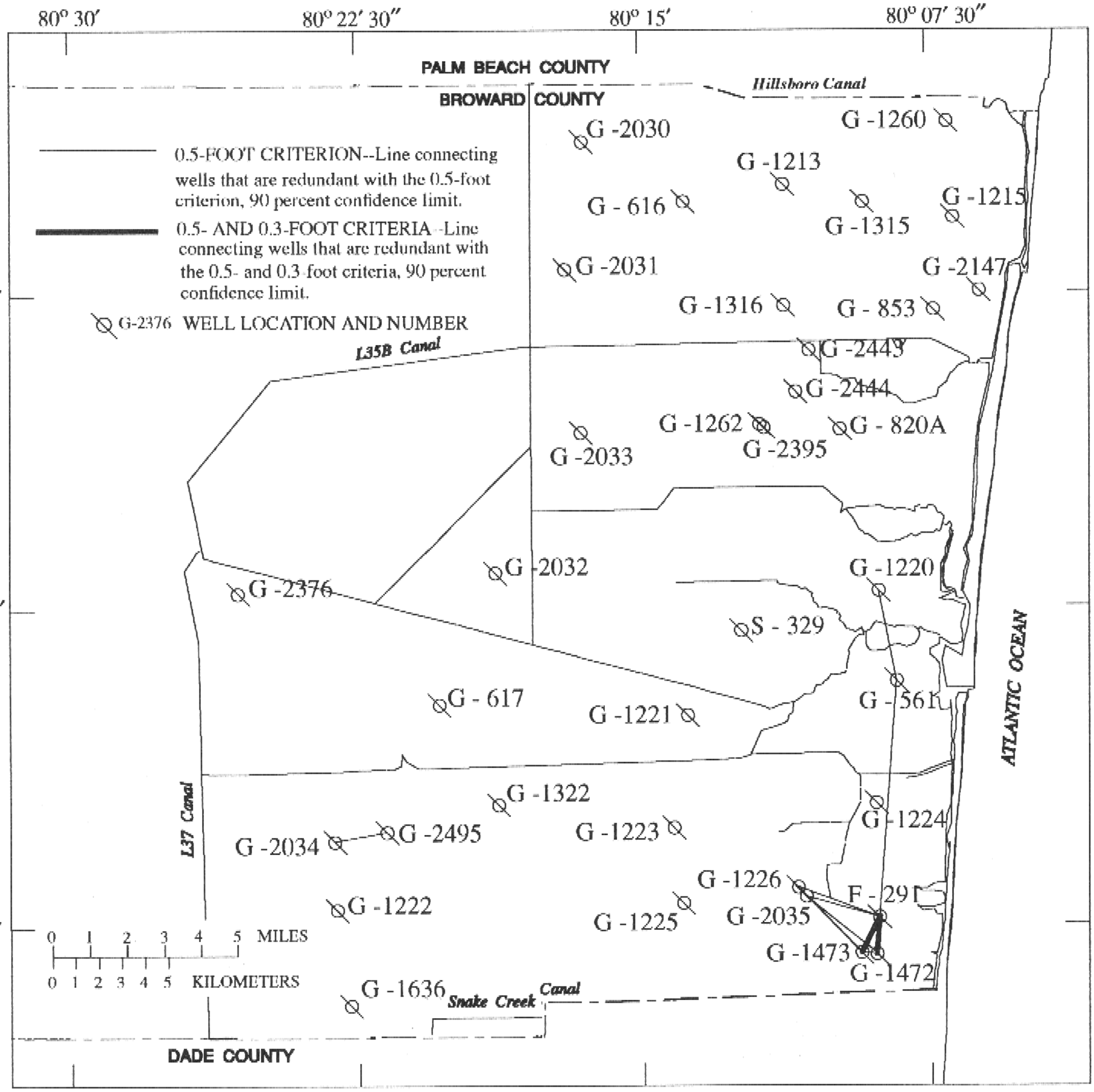


Figure 11. Pairs of redundant wells at 90 percent confidence limit using all data (0.5-foot and 0.3-foot criteria).

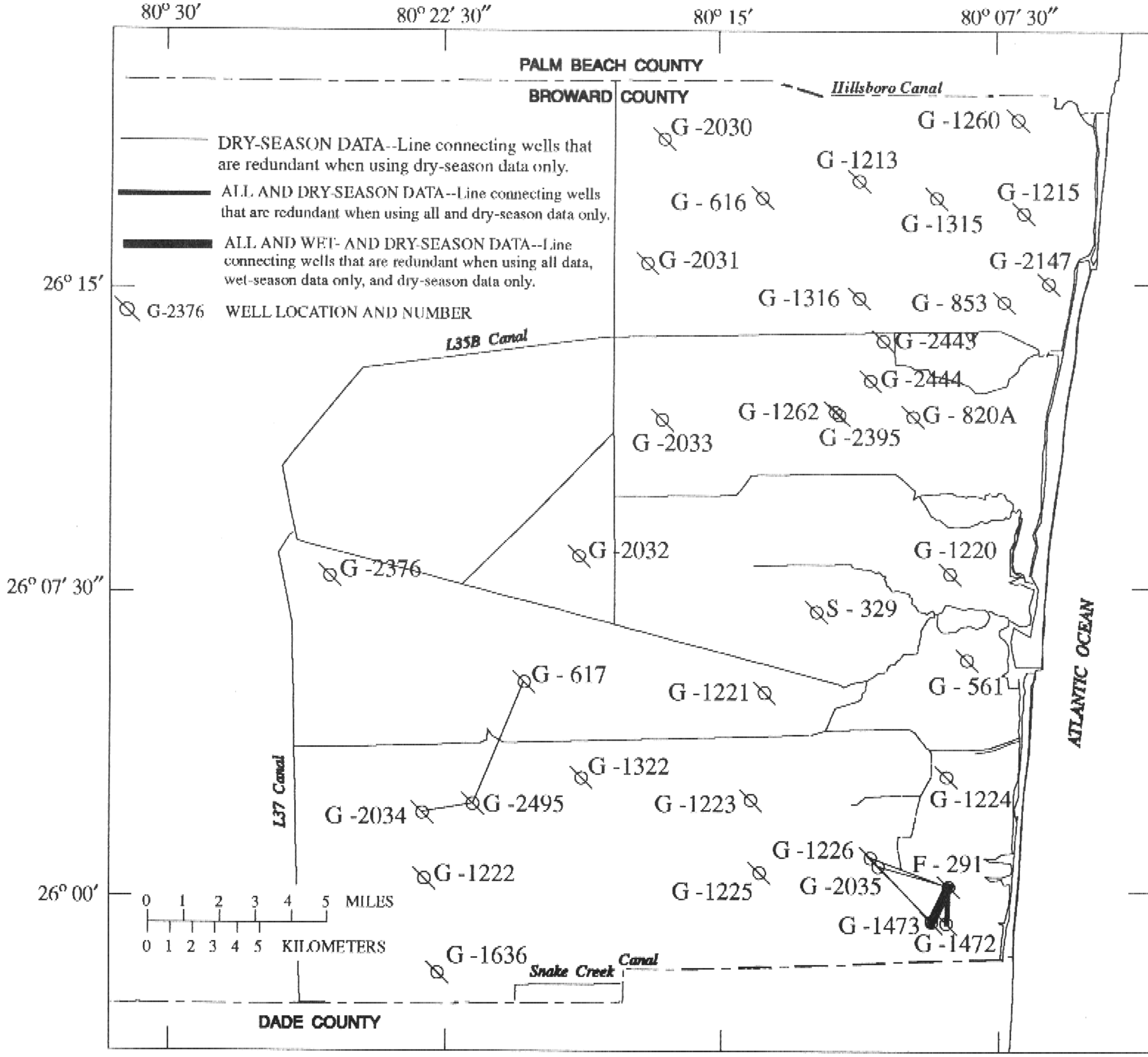


Figure 12. Pairs of redundant wells with 0.3-foot criterion at 90 percent confidence limit using all data, wet-season data only, and dry-season data only.

Data Verification

The potential sources of errors and estimations in the method are apparent because of the potential disadvantages discussed in the "*Spatial Method*" section. Therefore, data from a well not used in the development of the CPs are used for comparison with adjacent wells. If a CP of an adjacent well overlaps the new well, the data from the adjacent well should be able to be used to estimate the new well values within the given criteria.

There are many wells in Broward County that do not have continuous recorders, so only periodic water-level data are available from them (Haire and Lietz, 1991). The water-level data for any well can be expressed as perturbations about the mean by subtracting the mean of the data from each data value. These perturbations can be compared to the simultaneous perturbations of any other well to determine the percentage of time that they do not differ more than the given criteria. This result can be compared with the predicted accuracies from the CP analysis.

A reasonably large data set is needed to define perturbations about a mean. Well G-2409 (fig. 13), which has no continuous recorder, has one of the largest set of water-level values (110) between September 16, 1985, and September 25, 1992. There are significant differences between the data collected from well G-2409 and the continuous recorder wells used to calculate the CPs. The water levels recorded for well G-2409 are instantaneous at the time of measurement. The continuous recorder well values are daily maximums. If a rainfall event occurs after recording the water level at well G-2409 during the same day, the daily maximum might be significantly different than the value measured. Also, the limited number of data points at well G-2409 allows more error in a statistical analysis. Despite these limitations, well G-2409 presents an adequate set of noncontinuous recorder well data for verifying the CPs.

The CPs of the eight wells near well G-2409 are shown in figures 13 and 14. Well G-2409 with the 0.5-ft criterion at 90 percent confidence is within the CPs of wells F-291, G-1226, G-1472, G-1473, and G-2035 (fig. 13). Well G-2409 with the 0.3-ft criterion at 90 percent confidence is within the CPs of wells F-291, G-1472, and G-1743 (fig. 14). If the CPs are accurate, the data from these wells could be used to predict the perturbations at well G-2409 within these criteria.

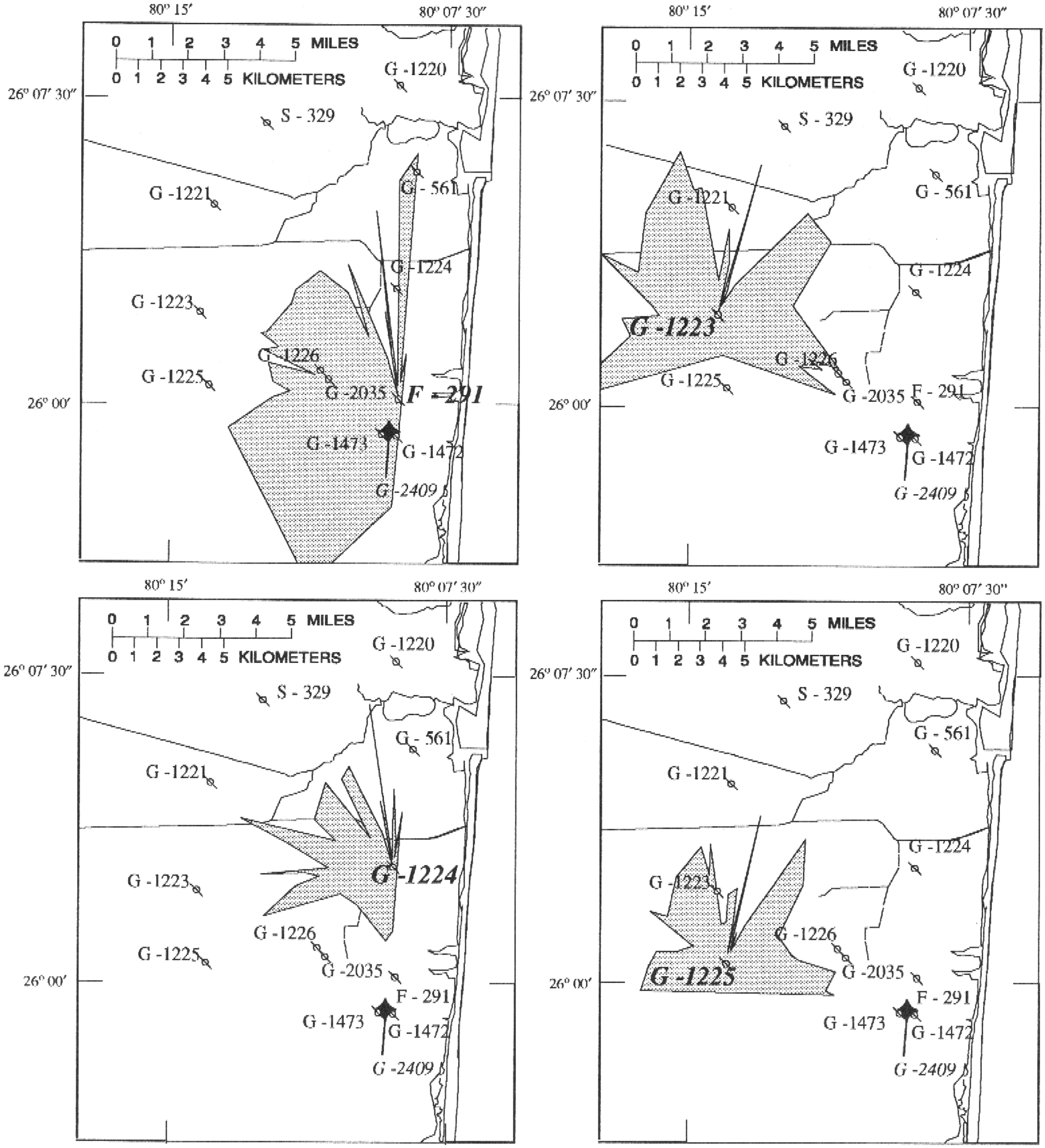
The water-level perturbations from the mean at well G-2409 were compared to those at the continuous recorder wells. The number of data points collected in which the perturbations at each continuous recorder well were within the 0.3- or 0.5-ft criteria of the perturbations at well G-2409 were calculated as a percentage of the total number of data points. This calculation was used to indicate the percentage of data collected at each of the wells that can be used to predict the water levels at well G-2409 within 0.5 ft as shown in figure 15. On the basis of the previously mentioned CPs, the amount of data at well G-2409 predicted by wells F-291, G-1226, G-1472, G-1473, and G-2035 was greater than 90 percent, indicating that the CPs are defined properly. However, the data at wells G-1224 and G-1225 can also be used to predict greater than 90 percent of the data at well G-2409 (95.3 percent and 90.7 percent, respectively) when the CPs do not make this indication. Both wells have CPs close to, but not overlapping, well G-2409, indicating that the CPs might be underestimated in some cases (fig. 13).

The percentage of data collected at each of the wells can be used to predict the water levels at well G-2409 within 0.3 ft (fig. 16). The CPs at wells F-291, G-1472, and G-1473 (fig. 14) indicate that the predicted level should be greater than 90 percent at these wells as shown in figure 16. The CPs also indicate that data from no other wells should be used to predict water levels at well G-2409 within 0.3 ft greater than 90 percent of the time as shown in figure 16. No errors were detected in the CP value for these criteria.

TEMPORAL ANALYSIS

The calculated autovariograms for each well for the period of record are used for temporal analysis (fig. 17). The entire period of record was used for each of the calculations. The approach to temporal analysis was to calculate the functions with a maximum lag time τ of $\sqrt{N} + 10$ where N is the number of days of data. When the maximum lag exceeds this fraction of the total data points, the functions become unreliable because of insufficient randomization. Thus, wells with short periods of record have short autovariogram functions. However, maximum time lags for all wells are high enough to calculate temporal confidence intervals.

The variogram criterion (γ_d) for 0.5-ft accuracy at 90 percent confidence was previously calculated by



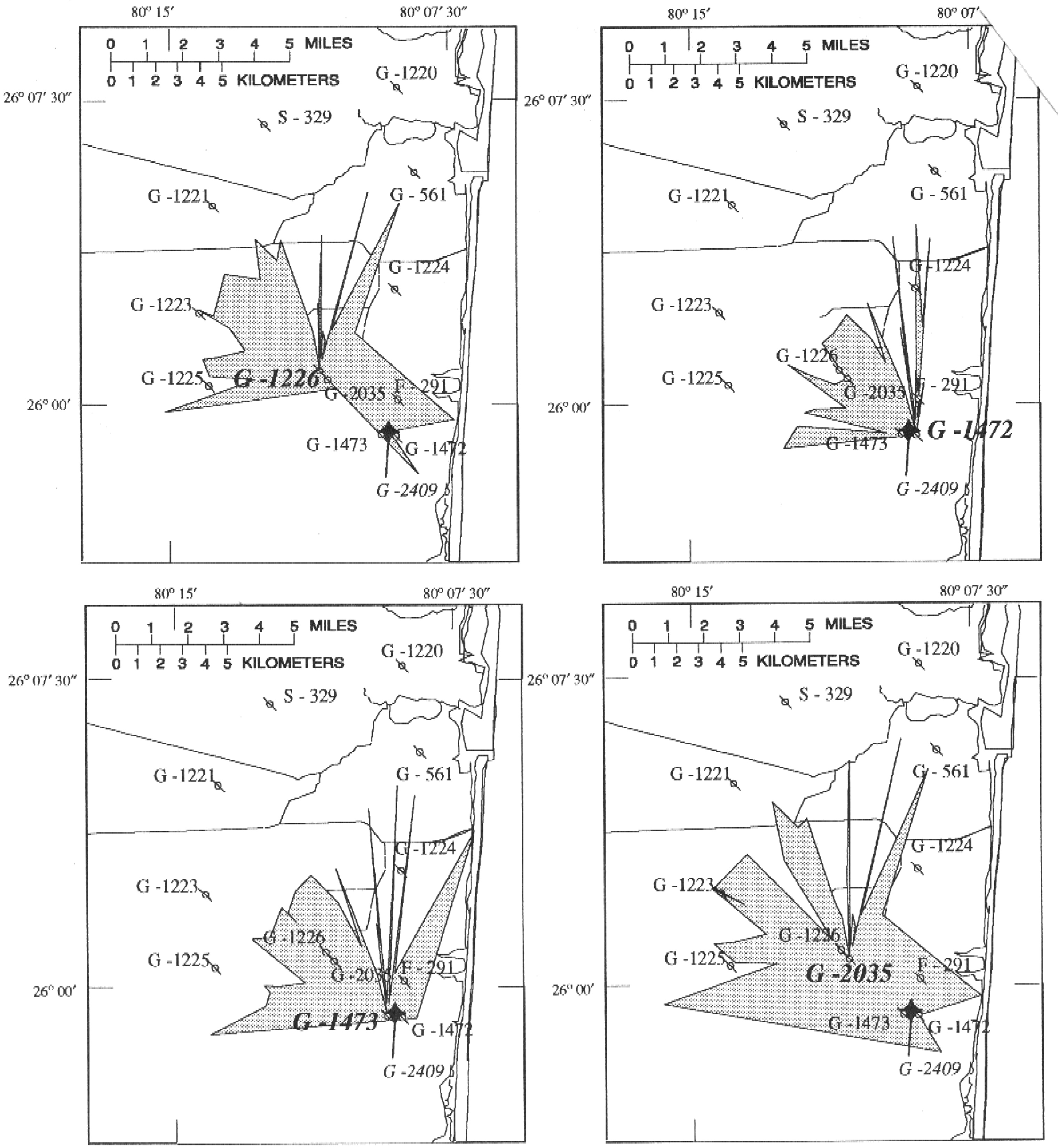
EXPLANATION

F - 291 CONFIDENCE POLYGON AND WELL NUMBER

G-1225 WELL LOCATION AND NUMBER

G-2409 COMPARISON WELL FOR VERIFICATION

Figure 13. Location of well G-2409 and confidence polygons of nearby wells for 0.5-foot criterion, 90 percent confidence limit.



EXPLANATION

F - 291 CONFIDENCE POLYGON AND WELL NUMBER

G-1225 WELL LOCATION AND NUMBER

G-2409 COMPARISON WELL FOR VERIFICATION

Figure 13. Location of well G-2409 and confidence polygons of nearby wells for 0.5-foot criterion, 90 percent confidence limit--Continued.

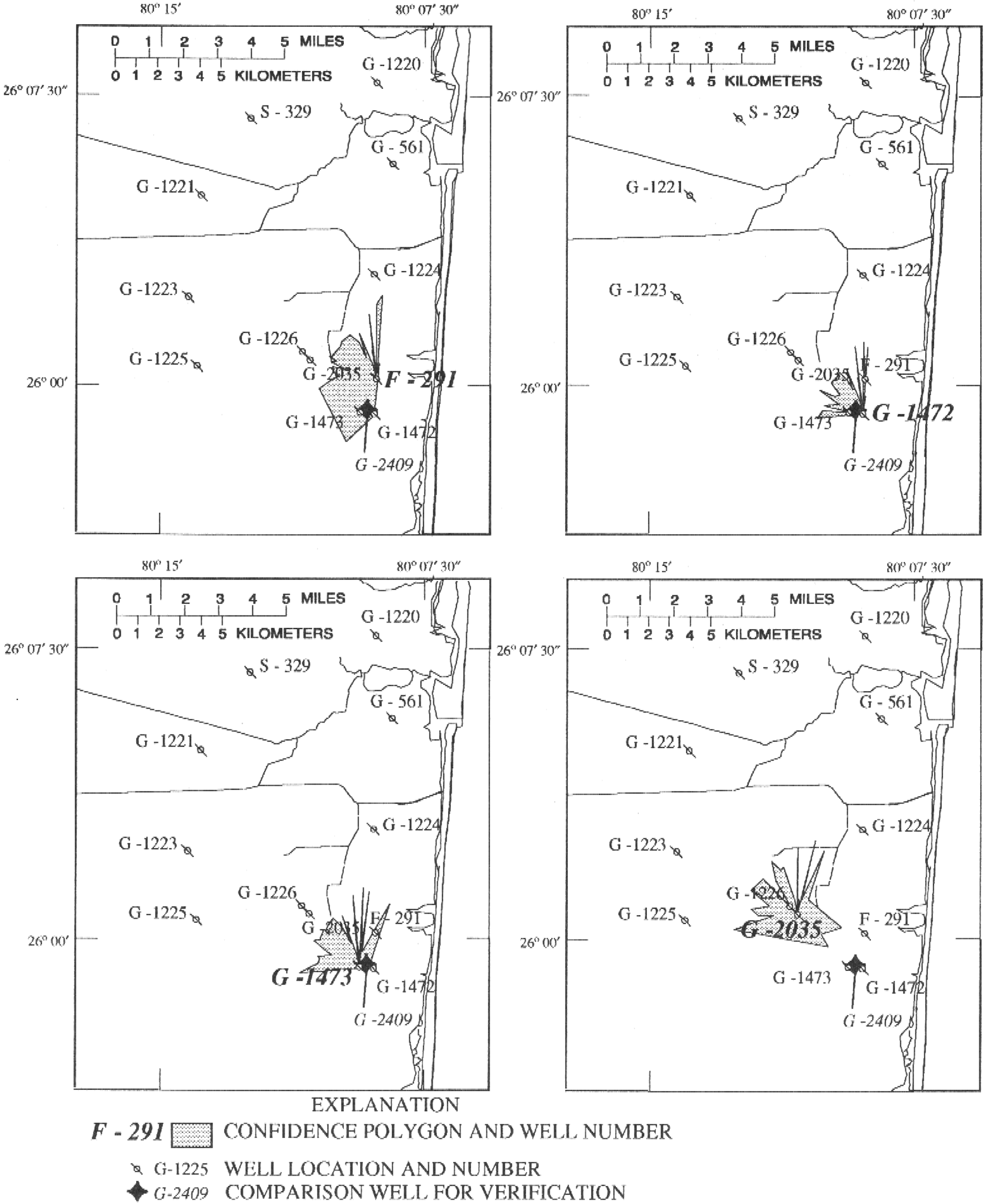


Figure 14. Location of well G-2409 and confidence polygons of nearby wells for 0.3-foot criterion, 90 percent confidence limit.

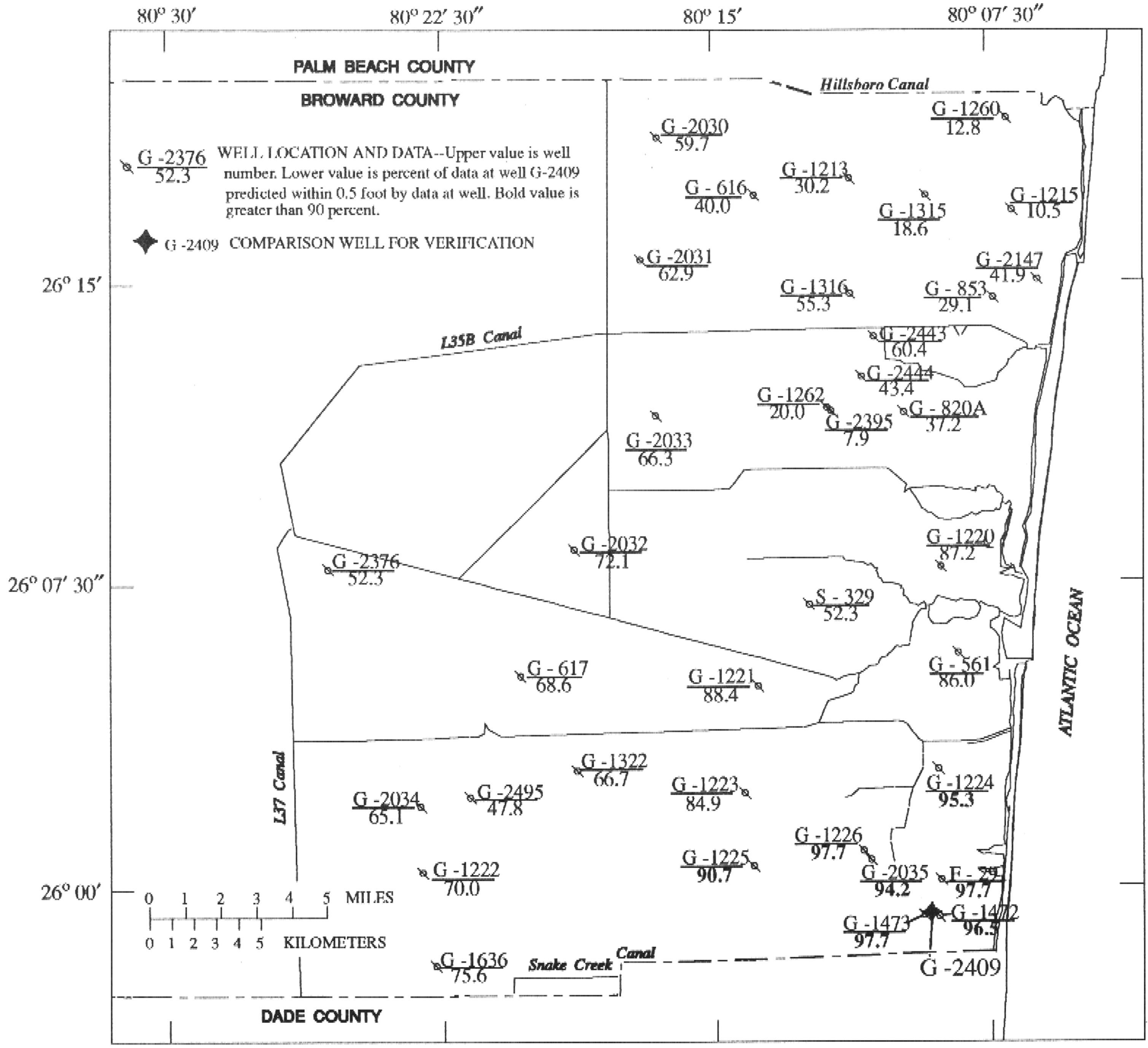


Figure 15. Percentage of data at well G-2409 predicted within 0.5 foot by data at each well.

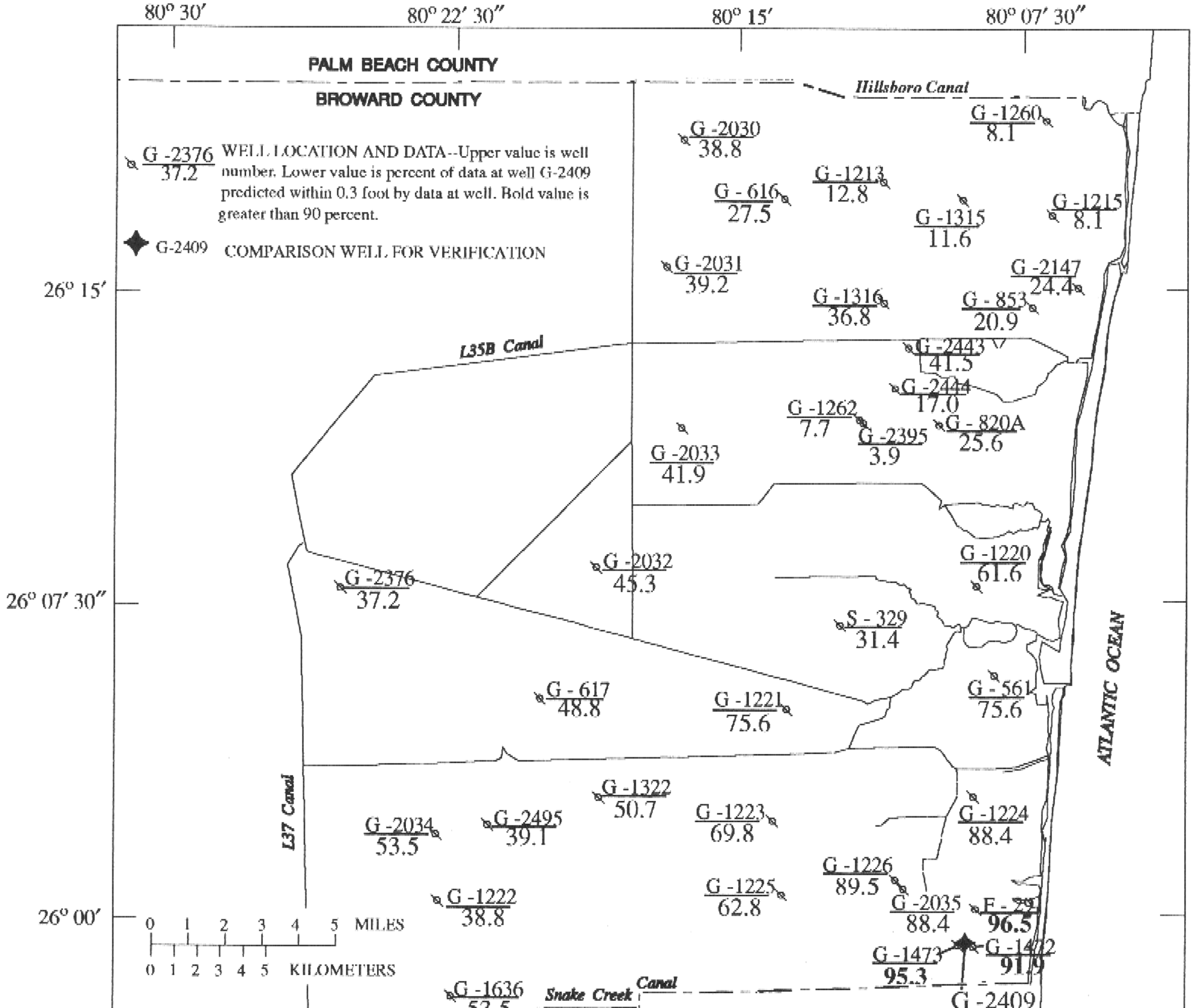


Figure 16. Percentage of data at well G-2409 predicted within 0.3 foot by data at each well.

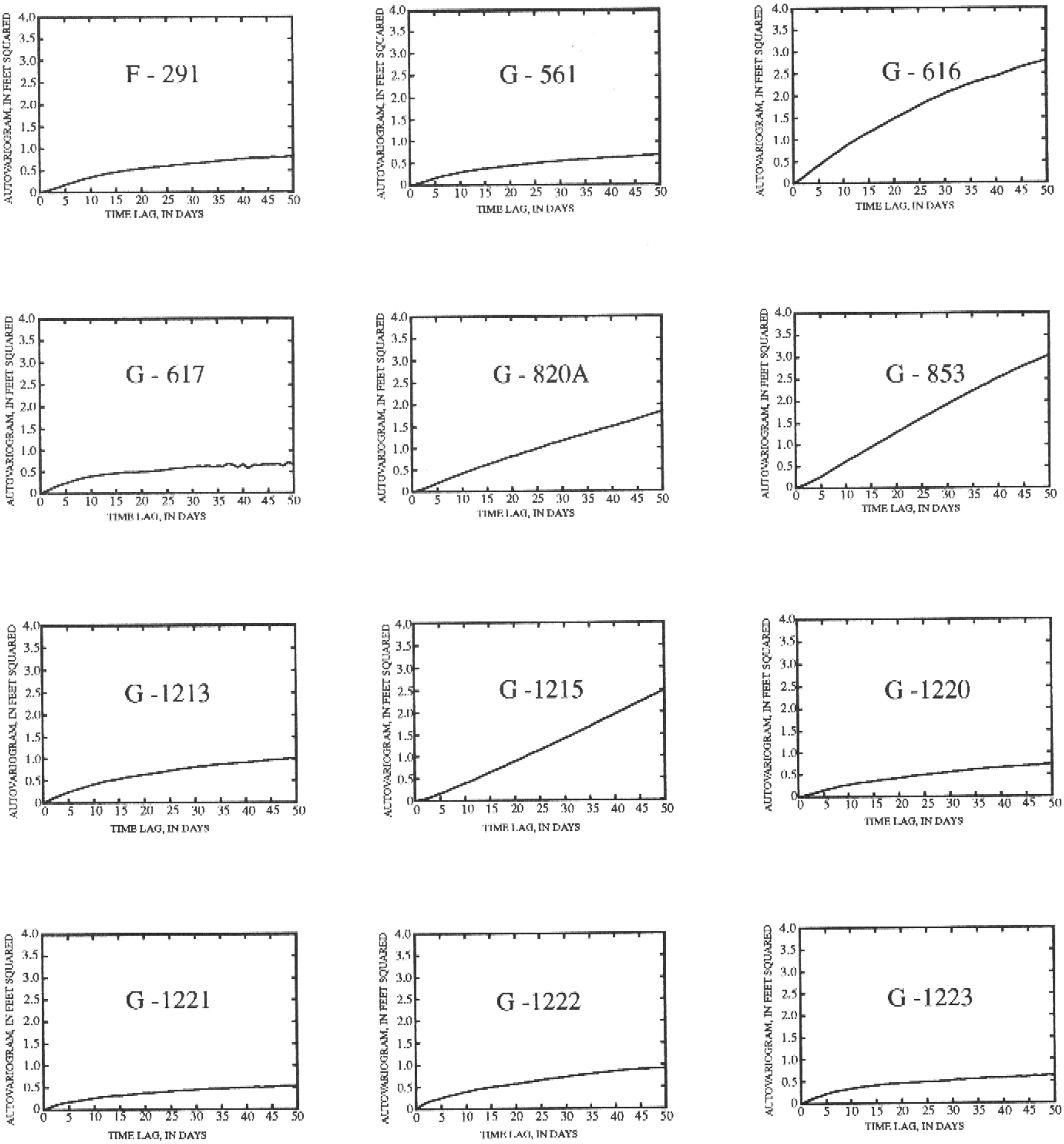


Figure 17. Calculated autovariograms for wells in the Broward County ground-water level monitoring network. Well locations shown in figure 2.

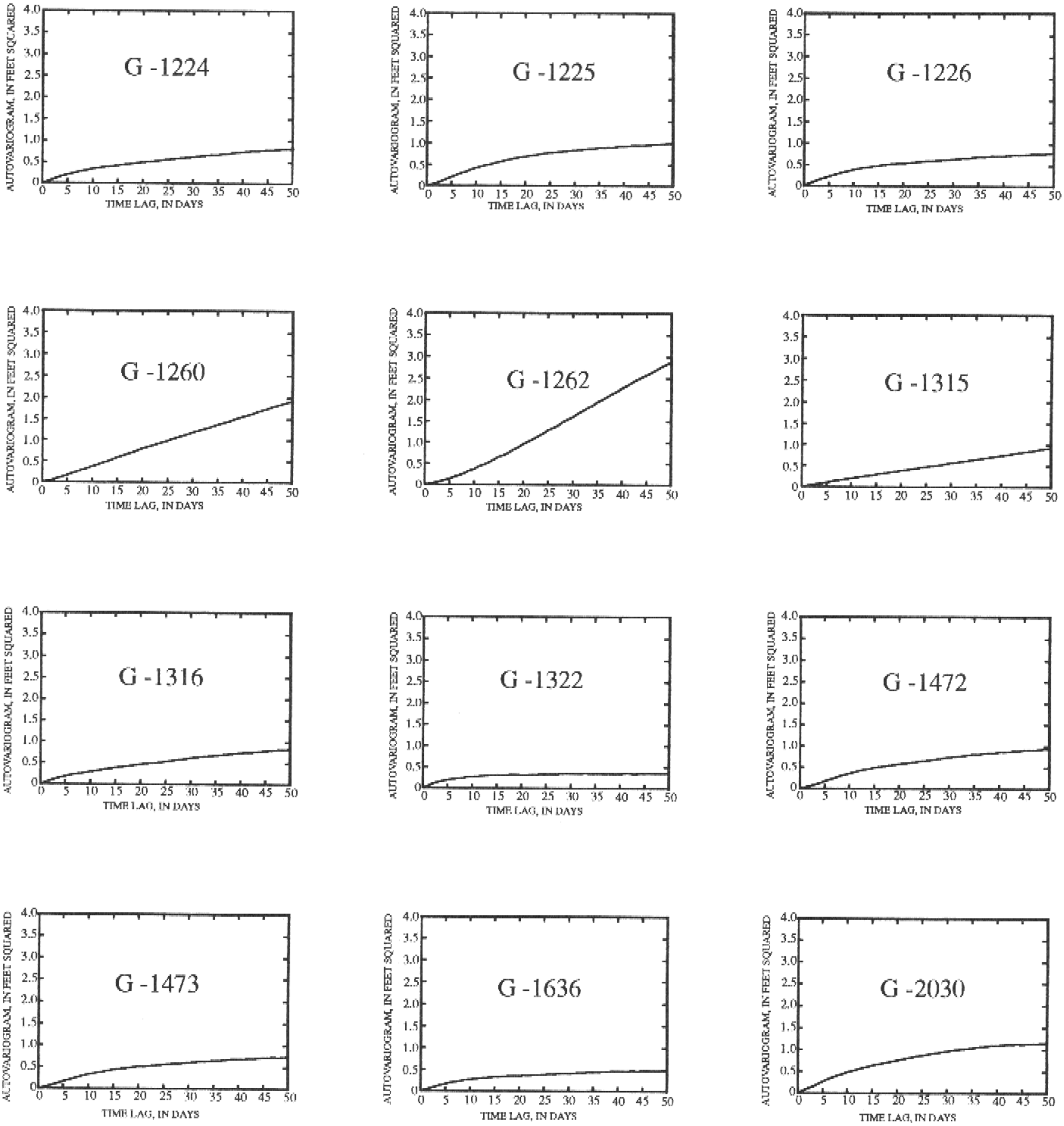


Figure 17. Calculated autovariograms for wells in the Broward County ground-water level monitoring network--Continued. Well locations shown in figure 2.

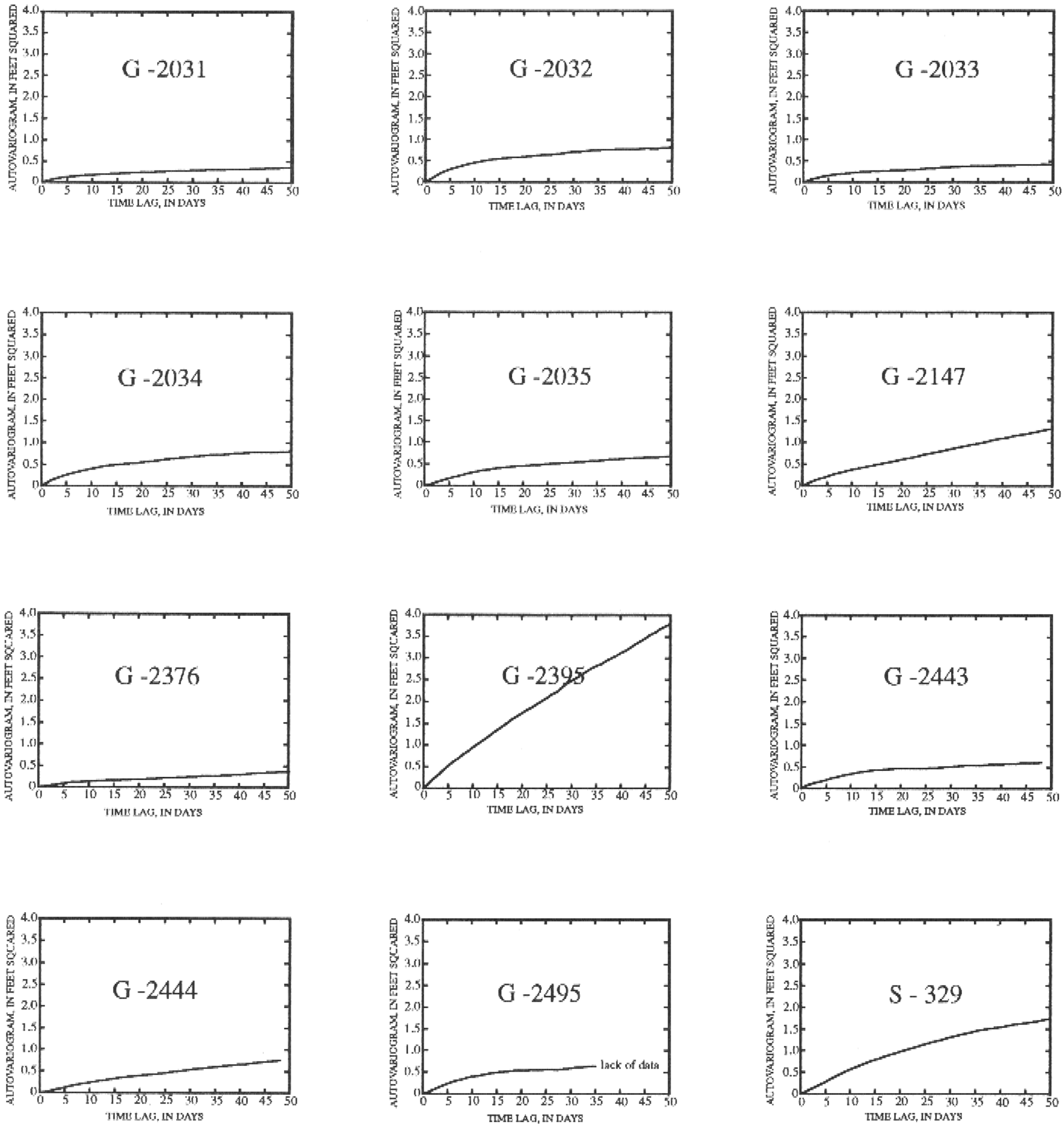


Figure 17. Calculated autovariograms for wells in the Broward County ground-water level monitoring network--Continued. Well locations shown in figure 2.

equation 17 with $\varepsilon = 0.5$ ft and $Z = 1.6448$ to be 0.0924 ft^2. The points on each plot in figure 17 where the autovariogram exceeds this variogram criterion were calculated and are given in table 5. Fractional daily values are linearly interpolated between daily values. The temporal confidence intervals range from 0.89 day (21 hours) at well G-2395 to 5.47 days at well G-2376.

Table 5. Temporal confidence intervals for 0.3 and 0.5 feet at 90 percent confidence limit

Well number	Temporal confidence interval for 0.3 foot (days)	Temporal confidence interval for 0.5 foot (days)
F-291	1.32	2.86
G-561	1.29	2.93
G-616	.54	1.36
G-617	.72	1.76
G-820A	1.29	2.74
G-853	.67	1.87
G-1213	.76	1.83
G-1215	1.48	3.39
G-1220	1.33	3.29
G-1221	.71	2.05
G-1222	.38	1.09
G-1223	.77	1.91
G-1224	.85	2.11
G-1225	1.16	2.50
G-1226	.62	1.79
G-1260	1.41	3.14
G-1262	1.63	3.55
G-1315	1.53	4.39
G-1316	.94	2.39
G-1322	.64	1.72
G-1472	1.38	3.00
G-1473	1.36	3.01
G-1636	1.02	2.59
G-2030	.80	1.88
G-2031	.92	2.65
G-2032	.49	1.32
G-2033	.92	2.49
G-2034	.53	1.44
G-2035	1.18	2.72
G-2147	.68	1.83
G-2376	2.02	5.47
G-2395	.32	.89
G-2443	.62	1.80
G-2444	1.71	3.91
G-2495	.82	2.00
S-329	.70	1.81

Because the temporal confidence intervals indicate a period of time over which water levels can be predicted from a single measurement, they correspond to maximum time intervals for an optimal measurement scheme. These temporal confidence intervals are short compared with any reasonable manual measurement protocol, and for most wells, water levels change so much within several days that predictions cannot be made with accuracy (table 5).

Well G-2395 has the only temporal confidence interval less than 1 day for 0.5 ft. This indicates that for the first lag point (1 day), the autovariogram in figure 17 was greater than the criterion (0.0924 ft^2). Because the field data are in the daily format, the existing data are insufficient to predict values at the well during the day with the desired accuracy. However, the raw field data are collected hourly.

The time lags at which the autovariograms exceed the criterion for 0.3-ft accuracy at 90 percent confidence were calculated and are given as temporal confidence intervals in table 5. As expected, the intervals are short, many less than 1 day.

NETWORK STATISTICS

The STARE coefficient development described in a previous section is applied to the network data (tables 3-5), and the results are reported for each well as adjusted STARE coefficients (table 6). The total area covered by the network (shaded area in fig. 6) is calculated by GIS to be 208.13 mi^2. The CP areas are normalized by this number. The temporal confidence intervals are normalized by the measurement interval of 1 day. The adjusted STARE coefficients and the well locations are shown in figure 18.

Well G-1472 is ranked as the least important well to the network with an adjusted STARE coefficient of 0.00 (table 6), which is primarily because of its redundancy. Well G-1223 is ranked as the most important well, with an adjusted STARE coefficient of 12.92 (table 6). This well has a relatively large CP area, no redundancy, and an average temporal confidence interval.

The STARE coefficient only indicates the significance of each well with respect to the entire network. No site-specific concerns are considered, and "holes" in the network are not specifically addressed by the STARE coefficients but rather by the map of CP

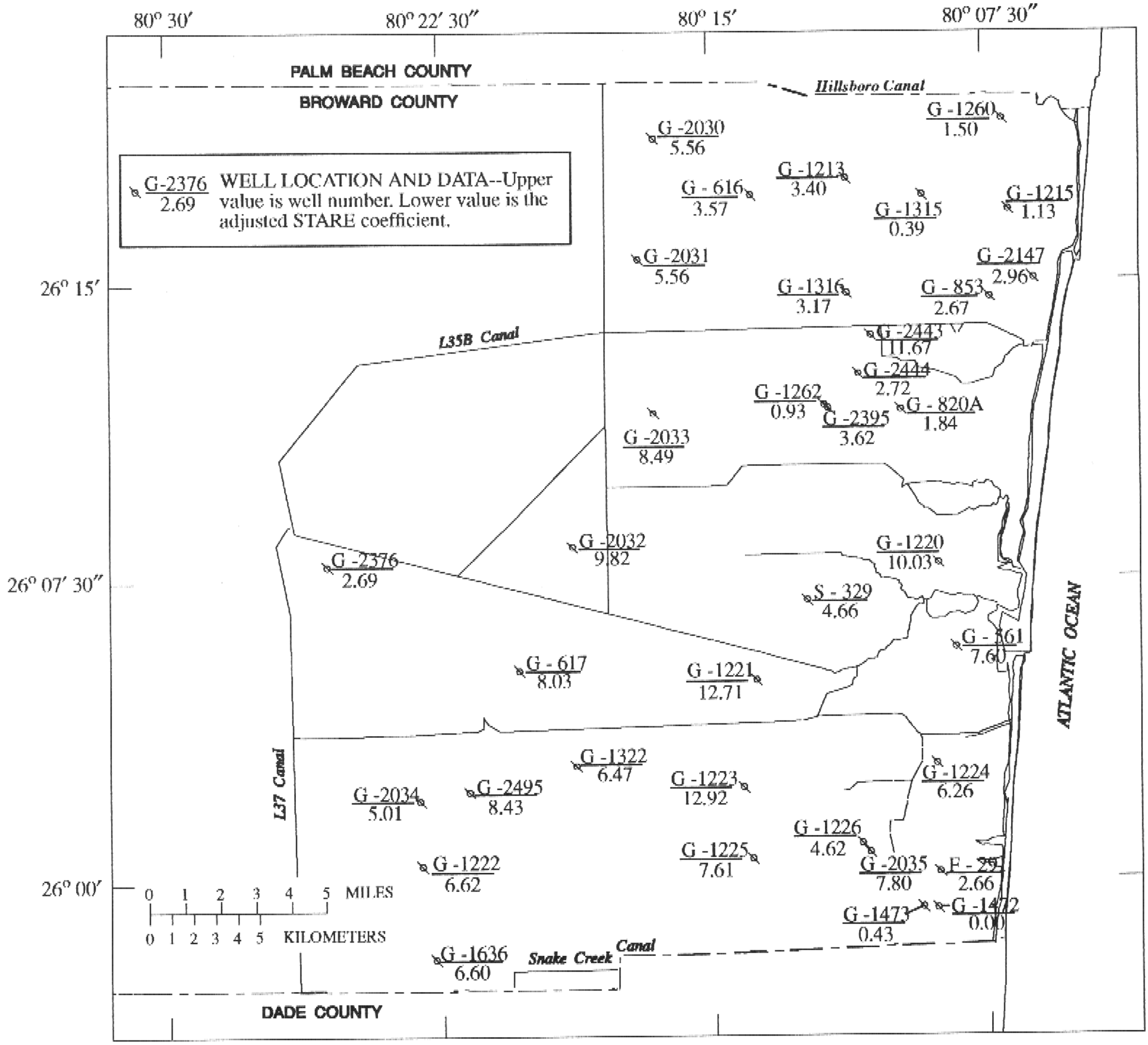


Figure 18. Adjusted Spatial and Temporal Adequacy and Redundancy Evaluation (STARE) coefficients for each well in the Broward County ground-water level monitoring network.

Table 6. Development of Spatial and Temporal Adequacy and Redundancy Evaluation (STARE) coefficients

Well number	Weighted confidence polygon area[1]	Redundancy[2]	Weighted temporal confidence interval[3]	Unadjusted STARE coefficient[4]	Adjusted STARE coefficient[5]
G-1472	3.53	5	3.00	-4.47	0.00
G-1315	.31	0	4.39	-4.08	.39
G-1473	6.97	8	3.01	-4.04	.43
G-1262	.01	0	3.55	-3.54	.93
G-1215	.05	0	3.39	-3.34	1.13
G-1260	.17	0	3.14	-2.97	1.50
G-820A	.11	0	2.74	-2.63	1.84
F-291	13.05	12	2.86	-1.81	2.66
G-853	.07	0	1.87	-1.80	2.67
G-2376	3.69	0	5.47	-1.78	2.69
G-2444	2.16	0	3.91	-1.75	2.72
G-2147	.32	0	1.83	-1.51	2.96
G-1316	1.09	0	2.39	-1.30	3.17
G-1213	.76	0	1.83	-1.07	3.40
G-616	.46	0	1.36	-.90	3.57
G-2395	.04	0	.89	-.85	3.62
G-1226	7.94	6	1.79	.15	4.62
S-329	2.00	0	1.81	.19	4.66
G-2034	3.98	2	1.44	.54	5.01
G-2031	3.74	0	2.65	1.09	5.56
G-2030	2.97	0	1.88	1.09	5.56
G-1224	3.90	0	2.11	1.79	6.26
G-1322	3.72	0	1.72	2.00	6.47
G-1636	4.72	0	2.59	2.13	6.60
G-1222	3.24	0	1.09	2.15	6.62
G-561	8.06	2	2.93	3.13	7.60
G-1225	5.64	0	2.50	3.14	7.61
G-2035	12.05	6	2.72	3.33	7.80
G-617	6.32	1	1.76	3.56	8.03
G-2495	8.96	3	2.00	3.96	8.43
G-2033	6.51	0	2.49	4.02	8.49
G-2032	6.67	0	1.32	5.35	9.82
G-1220	9.85	1	3.29	5.56	10.03
G-2443	9.00	0	1.80	7.20	11.67
G-1221	10.29	0	2.05	8.24	12.71
G-1223	10.36	0	1.91	8.45	12.92

[1]Weighted confidence polygon area = $\frac{confidence\ polygon\ area}{total\ area\ covered\ by\ network} \times 100$.

[2]Redundancy = total number of times well is part of a redundant pair.

[3]Weighted temporal confidence interval = $\frac{temporal\ confidence\ interval}{sampling\ interval}$.

[4]Unadjusted STARE coefficient = $\frac{confidence\ polygon\ area}{total\ area\ covered\ by\ network} \times 100 - redundancy - \frac{temporal\ confidence\ interval}{sampling\ interval}$.

[5]Adjusted STARE coefficient = unadjusted STARE coefficient - lowest unadjusted STARE coefficient.

areas. This is indicated by the low STARE coefficients assigned to wells G-1315, G-1262, G-1215, and G-1260 (table 6 and fig. 18). The low ratings for these northern wells are primarily due to the small CP areas and not due to redundancy (all wells have zero redundancy). These wells contribute little to the network coverage, but some are the only continuous recording wells in this area. Thus, any data are preferable to no data, and wells in an area that have small STARE coefficients should probably be retained. However, a well with a small STARE coefficient in close proximity to wells with significantly higher STARE coefficients is a prime candidate for removal from the network.

Well G-1472, the least important well to the network (table 6), should be the first well considered for removal. However, because the CPs of every well depends on the correlation with every other well, all CPs will be altered by the removal of G-1472. Therefore, it is necessary to analyze the removal of well G-1472 and the effect on the network.

The CP areas for 0.5-ft criterion at 90 percent confidence were recalculated without well G-1472, and the results are shown in figure 19. By comparison with figure 6, it is apparent that many wells show little change in the CP area, with the most marked changes near well G-1472. The new total area covered by the network is 205.40 mi^2. Thus, there is a 2.73-mi^2 reduction in the areal coverage when removing well G-1472. The weighted CP areas, divided by 205.40 mi^2, and expressed as a percent are given in table 7.

Immediately east of well G-1473, more area appears to be covered by the network in figure 19 than in figure 6. This is because of the CP of well G-2035. The removal of well G-1472 increases network coverage in one area (east of well G-1473) and decreases it in other areas (north of well F-291 and south of well G-1473). The correlation between wells G-2035 and G-1472 indicated a lower CP radius than the correlation of well G-2035 with other nearby wells (F-291 and G-1473). Therefore, the removal of well G-1472 caused the CP of well G-1472 to have a somewhat larger radius to the east (see caution no. 2 in the *"Spatial Method"* section). However, as previously stated, the total areal coverage of the network has been reasonably reduced by 2.73 mi^2.

The recalculation of the CPs is unnecessary for the 0.3-ft criterion at 90 percent confidence or the wet- and dry-season only situations because the results are only used to calculate the redundancy. Because redundancy occurs only when a CP radii is longer than the distance between wells, the number of times each well is a part of a redundant pair can be determined from the previous redundancy data, removing all pairs that contain G-1472. Thus, using the 0.5-ft criterion at 90 percent confidence, the redundant well pairs are: F-291/G-561, F-291/G-1226, F-291/G-1473, F-291/G-2035, G-561/G-1220, G-1226/G-1473, G-1226/G-2035, G-1473/G-2035, and G-2034/G-2495. The well pair that is redundant with the 0.3-ft criterion at 90 percent confidence is F-291/G-1473. The well pair that is effectively redundant in the wet-season analysis only is F-291/G-1473. Well pairs that are redundant in the dry-season analysis only are: F-291/G-1226, F-291/G-1473, F-291/G-2035, G-617/G-2495, G-1226/G-1473, G-1226/G-2035, and G-2034/G-2495. The number of times each well is part of a redundant pair is given in table 7.

Temporal confidence intervals are unchanged by the removal of well G-1472, except for the absence of well G-1472 (table 7). The unadjusted and adjusted STARE coefficients were recalculated and are given in table 7.

Comparison of table 7 with table 6 indicates the effects of removing the lowest ranked well on the STARE coefficient ranking. Although in table 6, well G-1315 follows well G-1472 as the least important well, table 7 indicates that the removal of well G-1472 has made G-1473 the least important well. The redundancy of well G-1473 is the major cause for its new ranking. Also, when ranking wells for least importance, well F-291 is ranked 8th when well G-1472 is included in the network (table 6), but 15th when the well is removed (table 7). Without the overlapping monitoring area from well G-1472, well F-291 becomes more important to the network even though it has the highest redundancy. A similar occurrence with well G-2035 is apparent as the ranking for this well is 27th in table 6 and 30th in table 7.

At the high STARE coefficient end of the tables, the coefficients and ranking change little compared to the low end. Well G-1223 is still the most important well to the network. The adjusted STARE coefficient for this well has only changed from 12.92 in table 6 to 12.90 in table 7. This analysis indicates that after eliminating well G-1472, well G-1473 would be the next removal. The procedure could be repeated for each well.

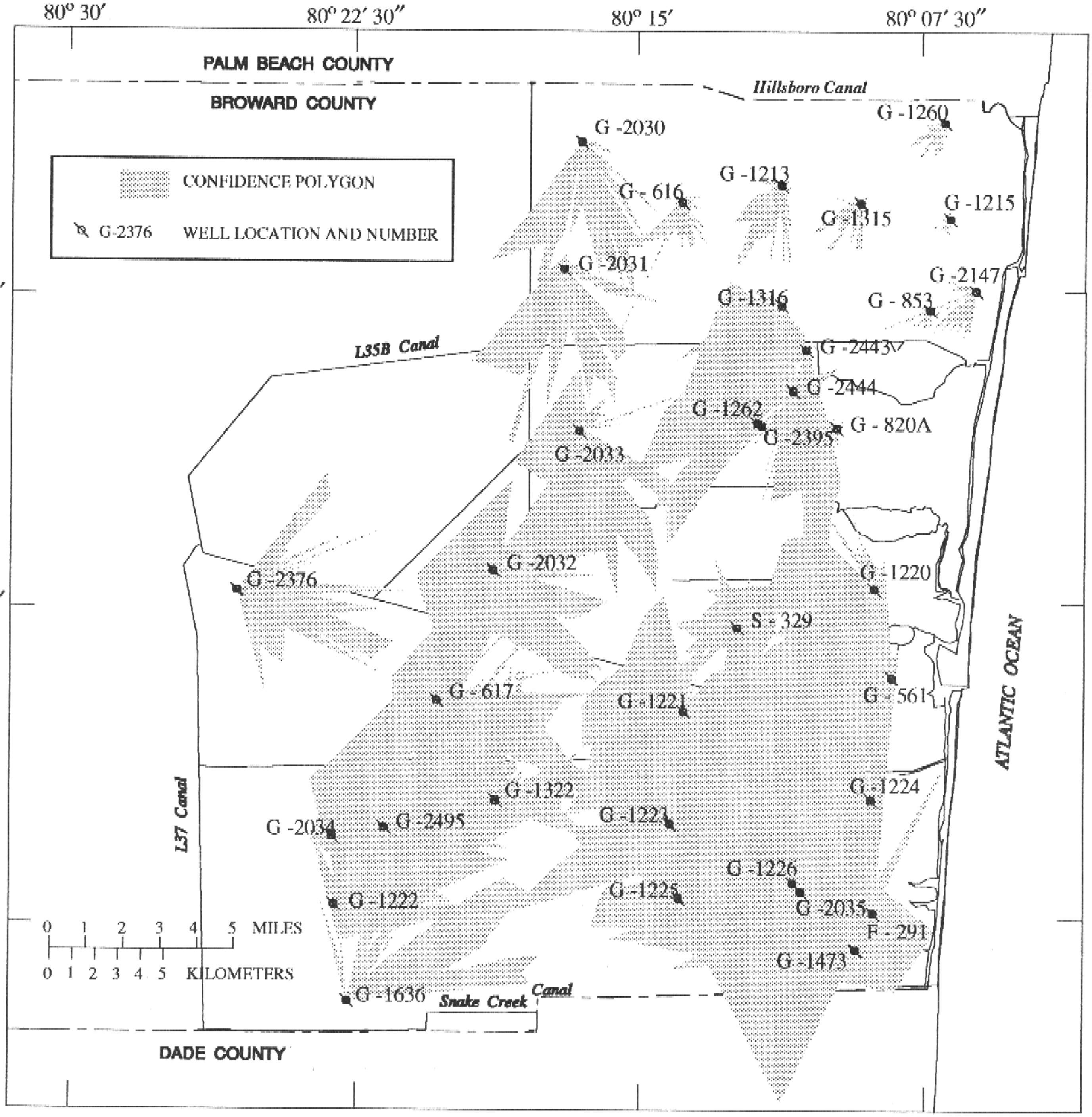


Figure 19. Confidence polygons for 0.5-foot criterion at 90 percent confidence limit with well G-1472 removed from the network.

Table 7. Recalculation of Spatial and Temporal Adequacy and Redundancy Evaluation (STARE) coefficients after removal of well G-1472

Well number	Weighted confidence polygon area[1]	Redundancy[2]	Weighted temporal confidence interval[3]	Unadjusted STARE coefficient[4]	Adjusted STARE coefficient[5]
G-1473	5.80	7	3.01	-4.21	0.00
G-1315	.31	0	4.39	-4.08	.13
G-1262	.01	0	3.55	-3.54	.67
G-1215	.05	0	3.39	-3.34	.87
G-1260	.17	0	3.14	-2.97	1.24
G-820A	.11	0	2.74	-2.63	1.58
G-853	.07	0	1.87	-1.80	2.41
G-2376	3.72	0	5.47	-1.75	2.46
G-2444	2.17	0	3.91	-1.74	2.47
G-2147	.32	0	1.83	-1.51	2.70
G-1316	1.10	0	2.39	-1.29	2.92
G-1213	.78	0	1.83	-1.05	3.16
G-616	.46	0	1.36	-.90	3.31
G-2395	.04	0	.89	-.85	3.36
F-291	11.52	9	2.86	-.34	3.87
S-329	2.02	0	1.81	.21	4.42
G-2034	4.03	2	1.44	.59	4.80
G-1226	8.64	6	1.79	.85	5.06
G-2031	3.76	0	2.65	1.11	5.32
G-2030	3.02	0	1.88	1.14	5.35
G-1224	3.95	0	2.11	1.84	6.05
G-1322	3.79	0	1.72	2.07	6.28
G-1636	4.76	0	2.59	2.17	6.38
G-1222	3.31	0	1.09	2.22	6.43
G-1225	5.73	0	2.50	3.23	7.44
G-561	8.19	2	2.93	3.26	7.47
G-617	6.45	1	1.76	3.69	7.90
G-2495	9.08	3	2.00	4.08	8.29
G-2033	6.62	0	2.49	4.13	8.34
G-2035	12.73	5	2.72	5.01	9.22
G-2032	6.77	0	1.32	5.45	9.66
G-1220	9.95	1	3.29	5.66	9.87
G-2443	9.10	0	1.80	7.30	11.51
G-1221	10.51	0	2.05	8.46	12.67
G-1223	10.60	0	1.91	8.69	12.90

[1]Weighted confidence polygon area = $\frac{confidence\ polygon\ area}{total\ area\ covered\ by\ network} \times 100$.

[2]Redundancy = total number of times well is part of a redundant pair.

[3]Weighted temporal confidence interval = $\frac{temporal\ confidence\ interval}{sampling\ interval}$.

[4]Unadjusted STARE coefficient = $\frac{confidence\ polygon\ area}{total\ area\ covered\ by\ network} \times 100 - redundancy - \frac{temporal\ confidence\ interval}{sampling\ interval}$.

[5]Adjusted STARE coefficient = unadjusted STARE coefficient - lowest unadjusted STARE coefficient.

SUMMARY AND CONCLUSIONS

A spatial and temporal analysis of continuous recorder ground-water network statistics was developed and performed on the ground-water level monitoring network in Broward County, Florida. This network was developed since 1940 with various wells being installed for different purposes. New canals and water-management structures, changing land use, and expansion of municipal well fields have resulted in wells no longer monitoring the situations for which they were originally designed. A need exists to evaluate the network and determine future needs and alternative monitoring network designs. Most of the wells have a period of record beginning in October 1973, so a sufficient data base exists for a spatial and temporal statistical analysis.

Considering the sparseness of the Broward County network, the spatial analysis technique was designed to develop polygons for each well that represent the areal coverage of each well. These "confidence polygons (CPs)" were defined as the boundary created by radial lines oriented toward the other wells. The lengths of these lines were determined as the statistically estimated distances to the points at which levels can be predicted within a certain confidence criterion.

The CPs represent an approximate area monitored by a well. However, the shape of any CP is affected by how many other wells are in the network. Therefore, the small variations in the boundary of a CP should not be given too much value. The wells at the outer edges of the network have no statistics exterior to the network for determination of CP radii, so the CPs of these wells tend to be truncated on their sides facing outward from the network. However, if the exterior wells are considered the boundary of the network, the CP area outside the network does not contribute to the network. The accuracy by which CPs truly represent the area monitored by the wells would be improved as the network is made more dense. Thus, the accuracy of the CPs is limited by the available data.

Analysis by CPs was applied to the Broward County network, using two different confidence criteria and wet- and dry-season data. Results indicated the effective coverage of the network and locations where data are unavailable. Well pairs with CPs that overlap well locations were defined as "effectively redundant." Comparison of predicted CP areas with an additional noncontinuous recorder well indicated that the polygons represent network coverage for the confidence criteria specified.

The temporal analysis technique considered time intervals as the temporal equivalent to the radial distances defined in the spatial analysis. Thus, based on the time varying data at each well, a time interval is determined beyond which water levels cannot be predicted within a given criterion. This "temporal confidence interval" at each well corresponds to the maximum measurement interval allowable for the desired confidence. Application of this technique to the Broward County network for two different criteria indicated that the temporal confidence interval is usually on the order of days, and at some wells for some criteria is less than the existing data interval (1 day).

The temporal analysis technique highlights the analogies and interdependence between the spatial and temporal properties of water levels measured at wells. As a monitoring well can be considered to only accurately predict water levels within a certain distance from the well (radius of the CP), it can also be considered to accurately predict water levels within a certain time interval after a measurement (temporal confidence interval). Both spatial and temporal considerations are important for accuracy in a water-level network.

A single coefficient reflecting both the spatial and temporal results was developed by combining all the results from the analyses. The STARE approach looks at three factors: the size of the CP, the number of times the well is part of a redundant pair, and the temporal confidence interval. The STARE coefficients for each well in the Broward County network were calculated. Wells with high redundancy or small CPs were at the bottom of the ranking. The effects of removing the well with the lowest STARE coefficient ranking were analyzed, and the STARE coefficient ranking for the remaining wells was calculated.

Although all attempts were made to construct the STARE coefficient to be as objective as possible, it has subjective characteristics. The size of the CP is affected by the accuracy criteria used, and the redundancy is affected by the accuracy criteria and the number of analyses considered. The accuracy criteria and the number of analyses considered must be selected such that CP size and redundancy are effective and proportional counterweights in the development of the STARE coefficient. This requires some trial and error to determine the best scheme. Additionally, the types of analyses considered (seasonal and criteria used) must be based on the type and accuracy of data required from the network.

REFERENCES CITED

Aller, L., Bennett, T., Lehr, J.H., and Petty, R.J., 1985, DRASTIC: A standardizing system for evaluating ground water pollution potential using hydrogeologic settings: Ada, Oklahoma, U.S. Environmental Protection Agency, 197 p.

Bardossy, A., and Bogardi, I., 1983, Network design for the spatial estimation of environmental variables: Applied Mathematical Computations, v. 12, p. 339-365.

Beyer, W.H., 1987, CRC standard mathematical tables: Boca Raton, Florida, CRC Press, Inc., 674 p.

Carrera, J., Usunoff, E., and Szidarovszky, F., 1984, A method for optimal observation network design for groundwater management: Journal of Hydrology, v. 73, p. 147-163.

Chirlin, G.R., and Dagan, G., 1980, Theoretical head variograms for steady flow in statistically homogeneous aquifers: Water Resources Research, v. 16, no. 6, p. 1001-1015.

Cooper, R.M., and Istok, J.D., 1988, Geostatistics applied to ground-water contamination, 1, methodology: Journal of Environmental Engineering, v . 114, no. 2, p. 270-285.

Dagan, G., 1989, Flow and transport in porous formations: New York, Springer-Verlag, 465 p.

Delhomme, J.P., 1978, Kriging in the hydrosciences: Advances in Water Resources, v. 1, no. 5, p. 251-266.

----- 1979, Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach: Water Resources Research, v. 15, no. 2, p. 269-280.

de Marsily, G., 1986, Quantitative hydrogeology: San Diego, California, Academic Press, 317 p.

Federal Register Notice, October 11, 1979, v. 44, no. 198.

Fish, J.E., 1988, Hydrogeology, aquifer characteristics, and ground-water flow of the surficial aquifer system, Broward County, Florida: U.S. Geological Survey Water-Resources Investigations Report 87-4034, 50 p.

Gambolati, G., and Volpi, G., 1979, Groundwater contour mapping in Venice by stochastic interpolators, 1, theory: Water Resources Research, v. 15, no. 2, p. 281-290.

Haire, W.J., and Lietz, A.C., 1991, Water resources data - Florida, 1991, volume 2B: south Florida ground water: U.S. Geological Survey Water-Data Report FL-91-2B, 491 p.

Herr, J.W., 1990, Application of drastic ground water pollution mapping methodology to the SFWMD: West Palm Beach, Florida, South Florida Water Management District, 93 p.

Journel, A., and Huijbregts, 1978, Mining geostatistics: San Diego, California, Academic Press, 280 p.

LeGrand, H.E., 1983, A standardizing system for evaluating waste-disposal sites: Worthington, Ohio, National Water Well Association, 72 p.

Lietz, A.C., 1992, Altitude of the water table in the Biscayne aquifer, Broward County, Florida: U.S. Geological Survey Open-File Report 92-31, 1 sheet.

Loaiciga, H.A., Charbeneau, R.J., Everett, L.G., and others, 1992, Review of ground-water quality monitoring network design: Journal of Hydraulic Engineering, v. 118, no. 1, p. 11-37.

McBratney, A.B., Webster, R., and Burgess, T.M., 1981, The design of optimum sampling schemes for local estimation and mapping of regionalized variables: Computational Geoscience, v. 7, no. 4, p. 331-334.

Olea, R.A. 1975, Optimum mapping techniques using regionalized variable theory, series on spatial analysis: Kansas Geological Survey Report No. 2, 137 p.

Seo, D.J., Krajewski, W.F., and Bowles, D.S., 1990, Stochastic interpolation of rainfall data from rain gages and radar using cokriging 1: Design of experiments: Water Resources Research, v. 26, no. 3, p. 469-477.

Shapiro, S.S., and Wilk, M.B., 1965, An analysis of variance test for normality (complete samples): Biometrika, v. 52, p. 591-611.

Snedecor, G.W., and Cochran, W.G., 1982, Statistical methods: Ames, Iowa State University Press, 507 p.

Statware Inc., 1990, Statit statistical reference manual: Corvallis, Oregon, Statware Inc., 247 p.

Yevjevich, V., 1982, Probability and statistics in hydrology: Littleton, Colorado, Water Resources Publications, 302 p.

APPENDIXES

APPENDIX 1. PROGRAMS AND FILES

The following programs and files are used in the analysis and are listed in the order they are used. The programs along with the input-output files are in bold characters.

SPATIAL ANALYSIS

list	ASCII text file containing the names of the field data files.
list1	ASCII text file containing the names of the well sites in the format recognized by GIS.
prepare	FORTRAN program writes Unix script to calculate basic statistics for all well pairs; program writes to file **implement**.
implement	Unix script, which, for each well pair, calls **sort** program to sort data, executes **run** to calculate statistical parameters, and outputs to **mainout.**
sort	FORTRAN program sorts data for one-to-one correspondence in dates. Two variants exist: **sortd** does the same for dry season data only and **sortw** for wet season.
run	Unix script calls STATIT package program CORR to calculate basic statistics and SWNORM to implement Shapiro-Wilk normality test.
mainout	ASCII text file of statistics output from **implement**. Not in convenient format.
outsort	FORTRAN program converts data in **mainout** to usable format in **cordata.**
cordata	ASCII text file contains basic statistics in the order: first well name, second well name, number of concurrent days, first well variance, second well variance, covariance, correlation coefficient, correlation significance t value, probability of significant correlation, R^2 value for normal distribution test, and R^2 value for lognormal distribution test.
distcheck	FORTRAN program uses data in **cordata** to calculate best and worst fits for the normal and lognormal distributions. Output goes to terminal screen.
maxmin	FORTRAN program uses data in **cordata** to calculate maximum and minimum values of number of concurrent days, variance, covariance, and correlation coefficient. Output goes to file **maxminout.** Prints to terminal screen values and wells without significant correlations.
vario	FORTRAN program uses user input head accuracy with data in **cordata** to calculate variogram for each well pair, percent of distance between wells to point of desired accuracy for spherical, exponential, and Gaussian semivariogram functions. Output data in this order to file **variout.**
varioplot	FORTRAN program uses data from **variout** to display variograms from each well to all others. Data for scatter plots (as in fig. 7) put out to **fort.(well number)** files.
eqstats	FORTRAN program uses data in **variout** to calculate average differences between spherical-exponential equations and between exponential-Gaussian equations in calculated CP radii. The same comparisons are expressed as a percent. Results are sent to terminal screen.
verif	FORTRAN program uses data from files named in user-defined input file to compare with data in the file **table.2409.** The name of the file that contains the input file names and criteria are specified by the user when executing. The deviations between the data sets are analyzed, and the percent of data that is within the criteria is output to the screen.

TEMPORAL ANALYSIS

chk.com	Unix script puts field data files into **(well name).chk** files without headers or missing data points.
auto.com	Unix script executed in STATIT. Using data in **(well name).chk** files, the autocorrelation and autocovariance files are calculated. Output data goes to **auto.dat.**
autovar	FORTRAN program takes user input autovariogram criteria and data in **auto.dat.** Autovariograms are calculated for each well (put out to **fort.(well number)** files) and maximum measurement intervals (put out to **autovar.out**).

PREPARE PROGRAM

```
C      FORTRAN program writes Unix script to calculate basic
C      statistics for all well pairs, program writes to file
C      implement.
C
       character x*9
       dimension x(37)
       open (unit=20,file='list')
       do 10 i=1,1000
       read(20,100,end=11) X(I)
       imax=i
 10    continue
 11    continue
 100   format(a9)
       close (20)
       open (unit=21,file='implement')
       write(21,*) 'ln -s log fort.16'
       do 20 i=1,imax-1
       do 30 j=i+1,imax
       write(21,*) 'rm o1'
       write(21,*) 'rm o2'
       write(21,*) 'rm output'
       write(21,*) 'ln -s ',x(i),' fort.11'
       write(21,*) 'ln -s ',x(j),' fort.12'
       write(21,*) 'ln -s o1 fort.13'
       write(21,*) 'ln -s o2 fort.14'
       write(21,*) 'ln -s o3 fort.15'
       write(21,*) '/lhome/edswain/stat/sort'
       write(21,*) 'rm fort.11'
       write(21,*) 'rm fort.12'
       write(21,*) 'rm fort.13'
       write(21,*) 'rm fort.14'
       write(21,*) 'rm fort.15'
       write(21,*) 'statit<<*'
       write(21,*) 'exec run'
       write(21,200) 'quit'
       write(21,300) '*'
       write(21,*) 'echo ',x(i),' ',x(j),' >> mainout'
       write(21,*) 'cat output >> mainout'
 30    continue
 20    continue
       write(21,*) 'rm fort.??'
       close(21)
 200   format(a4)
 300   format(a1)
       end
```

SORT PROGRAM

```
C      THIS PROGRAM SORTS THE DATA IN TWO FILE SO THAT
C      THERE IS A 1 TO 1 CORRESPONDENCE IN DATES
C
       INTEGER YR1,MO1,DA1,DATE1,YR2,MO2,DA2,DATE2
       CHARACTER JUNK*29,WEL1*15,WEL2*15
       DO 5 I=1,2
       READ(11,*)
       READ(12,*)
 5     CONTINUE
       READ(11,300) JUNK,WEL1
       READ(12,300) JUNK,WEL2
       READ(11,*)
       READ(12,*)
       READ(11,*) YR1,MO1,DA1,H1
       READ(12,*) YR2,MO2,DA2,H2
 10    CONTINUE
       DATE1=10000*YR1+100*MO1+DA1
       DATE2=10000*YR2+100*MO2+DA2
       IF(DATE1.EQ.DATE2) GO TO 20
       IF(DATE1.GT.DATE2) THEN
       READ(12,*) YR2,MO2,DA2,H2
       GO TO 10
       ELSE
       READ(11,*) YR1,MO1,DA1,H1
       GO TO 10
       ENDIF
 20    CONTINUE
       IF(H1.GT.-1000.AND.H2.GT.-1000) THEN
       WRITE(13,200) DATE1,H1
       WRITE(14,200) DATE2,H2
       ENDIF
       READ(11,*,END=99) YR1,MO1,DA1,H1
       READ(12,*,END=99) YR2,MO2,DA2,H2
       DATE1=10000*YR1+100*MO1+DA1
       DATE2=10000*YR2+100*MO2+DA2
       GO TO 20
 99    CONTINUE
       REWIND (13)
       REWIND (14)
       DO 30 I=1,ICOUNT
       READ(13,200) DATE1,H1
       READ(14,200) DATE2,H2
       HDEV=H1-H2
       HDEVL=LOG(H1-H2+100.)
       WRITE(15,400) HDEV,HDEVL
 30    CONTINUE
       DO 40 I=1,99999
       READ(16,*,END=50)
 40    CONTINUE
 50    CONTINUE
       WRITE(16,*) WEL1,WEL2,ICOUNT
 100   FORMAT(I4,I11,I13,F21.0)
 200   FORMAT(I10,F16.2)
 300   FORMAT(A29,A15)
 400   FORMAT(2F16.5)
       END
```

RUN SCRIPT

```
/* Unix script calls STATIT package program CORR, */
/* to calculate basis statistics, and SWNORM to */
/* implement Shapiro-Wilk normality test. */
clear
log output
read o1 into a,x1 /CASES = 9999
read o2 into d,x2 /CASES = 9999
read o3 into x3,x4 /CASES = 1000
corr x1,x2 /COV /TEST
swnorm x3
swnorm x4
```

OUTSORT PROGRAM

```
C       FORTRAN program converts data in mainout
C       to usable format in cordata.
        character char1*19
        open(unit=10,file='mainout')
        open(unit=11,file='cordata')
 5      continue
        read(10,100,end=70) char1
        do 10 i=1,12
        read(10,*)
 10     continue
        read(10,200) ncases
        do 20 i=1,4
        read (10,*)
 20     continue
        read(10,300) var1
        read(10,400) covar,var2
        do 30 i=1,10
        read(10,*)
 30     continue
        read(10,500) corr,t,pt
        do 40 i=1,12
        read(10,*)
 40     continue
        read(10,600) rnorm
        do 50 i=1,37
        read(10,*)
 50     continue
        read(10,600) rlnorm
        write(11,700) char1,ncases,var1,var2,covar,corr,t,pt,
       1 rnorm,rlnorm
        do 60 i=1,26
        read(10,*,end=70)
 60     continue
        go to 5
 70     continue
        close(10)
        close(11)
 100    format(a19)
 200    format(46x,i4)
 300    format(26x,f9.0)
 400    format(26x,f9.0,1x,f9.0)
 500    format(25x,f7.0,7x,f6.0,16x,f7.0)
 600    format(44x,f6.0)
 700    format(a19,1x,i4,1x,8f9.5)
        end
```

DISTCHECK PROGRAM

```
C      FORTRAN program uses data in cordata to calculate
C      best and worst fits for the normal and lognormal
C      distributions. Output goes to terminal screen.

       open(unit=10,file='cordata')
       ic=0
       icl=0
       bestfn=-9.
       bestfl=-9.
       worsfn=9.
       worsfl=9.
       ave=0.
       avel=0.
       save=0.
       savel=0.
       icount=0
       do 10 i=1,9999999
       read(10,100,end=11) fit,fitl
       icount=icount+1
       ave=ave+fit
       avel=avel+fitl
       save=save+fit**2
       savel=savel+fit**2
       if(fit.gt.fitl) then
       ic=ic+1
       else
       icl=icl+1
       endif
       if(fit.gt.bestfn) bestfn=fit
       if(fit.lt.worsfn) worsfn=fit
       if(fitl.gt.bestfl) bestfl=fitl
       if(fitl.lt.worsfl) worsfl=fitl
 10    continue
 11    continue
       ave=ave/float(icount)
       avel=avel/float(icount)
       save=save/float(icount)
       savel=savel/float(icount)
       var=save-ave**2
       varl=savel-avel**2
       write(*,*) ic,' fitting normal distribution'
       write(*,*) icl,' fitting log-normal distribution'
       write(*,*) 'best fit for normal data = ',bestfn
       write(*,*) 'worst fit for normal data = ',worsfn
       write(*,*) 'best fit for log-normal data = ',bestfl
       write(*,*) 'worst fit for log-normal data = ',worsfl
       write(*,*) 'normal data mean = ',ave,' variance = ',var
       write(*,*) 'lognormal data mean = ',avel,
      1 ' variance = ',varl
 100   format(81x,f7.0,2x,f7.0)
       end
```

MAXMIN PROGRAM

```
C       FORTRAN program uses data in cordata to calculate
C       maximum and minimum values of; number of concurrent
C       days, variance, covariance, and correlation coefficient.
C       Output goes to file maxminout. Prints to terminal screen
C       values and wells without significant correlations.

        character char1*19
        open(unit=10,file='cordata')
        nmax=0
        nmin=999999
        varmax=-999999
        varmin=999999
        comax=-999999
        comin=999999
        cormax=-999999
        cormin=999999
        tmax=-999999
        tmin=999999
        ptmax=-999999
        ptmin=999999
        do 10 i=1,9999999
        read(10,100,end=11) char1,ncases,var1,var2,covar,corr,t,pt,
       1 rnorm,rlnorm
        if(ncases.gt.nmax) nmax=ncases
        if(ncases.lt.nmin) nmin=ncases
        if(var1.gt.varmax) varmax=var1
        if(var2.gt.varmax) varmax=var2
        if(var1.lt.varmin) varmin=var1
        if(var2.lt.varmin) varmin=var2
        if(covar.gt.comax) comax=covar
        if(covar.lt.comin) comin=covar
        if(corr.gt.cormax) cormax=corr
        if(corr.lt.cormin) cormin=corr
        if(t.gt.tmax) tmax=t
        if(t.lt.tmin) tmin=t
        if(pt.gt.ptmax) ptmax=pt
        if(pt.lt.ptmin) ptmin=pt
        if(pt.gt.0.10) write(*,*) char1,pt
  10    continue
  11    continue
        close (10)
        open(unit=11,file='maxminout')
        write(11,200) nmax,nmin,varmax,varmin,comax,comin,cormax,
       1 cormin,tmax,tmin,ptmax,ptmin
        close (11)
 100    format(a19,1x,i4,1x,8f9.5)
 200    format(i4,1x,i4,/,2f9.5,/,2f9.5,/,2f9.5,/,2f9.5,/,2f9.5)
        end
```

VARIO PROGRAM

```
C      FORTRAN program uses user input head accuracy with
C      data in cordata to calculate: variogram for each
C      well pair, percent of distance between wells to
C      point of desired accuracy for spherical,
C      exponential, and Gaussian semi-variogram functions.
C      Outputs data in this order to file variout.

       dimension wdat(50),wdis(50)
       character well*9,wel2*9,wel3*9,wel4*9,wdat*9,wdis*9
       open(unit=10,file='cordata')
       open(unit=20,file='variout')
       open(unit=25,file='variout1')
       open(unit=30,file='list')
       open(unit=40,file='list1')
       write(*,*) ' ENTER ACCURACY OF STAGE'
       read(*,*) acc
       gamma=((acc/1.6448)**2)/2.
       do 5 i=1,36
       read(30,200)wdat(i)
       read(40,200)wdis(i)
 5     continue
       close(30)
       close(40)
       do 10 i=1,9999999
       read(10,*,end=11) well,wel2,ncases,var1,var2,covar,
      1 corr,t,tprob
       var=(var1+var2)/2.
       semiva=var-covar
c
c      * a spherical semivariogram is tried *
c
       a=1.
 20    continue
       al=a
       a=(1.5*a**2-0.5)*var/semiva
       a=(abs(a))**(1./3.)
       if(abs((a-al)/a).gt.0.001) go to 20
       x1=1.
 30    continue
       x11=x1
       x1=1./(1.5/a-0.5*x1**2/a**3)*gamma/var
       if(abs((x1-x11)/x1).gt.0.001) go to 30
c
c      * an exponential semivariogram is tried *
c
       omgw=1.-semiva/var
       if(omgw.lt.0.000001) then
       x2=99999.
       go to 35
       endif
       a=-1./(log(omgw))
       x2=-a*log(1-gamma/var)
 35    continue
c
c      * a gaussian semivariogram is tried *
c
       omgw=1.-semiva/var
       omgw1=1.-gamma/var
       if(omgw.gt.0.99999.or.omgw1.gt.0.99999.or.
      1 omgw.lt.0.000001.or.omgw1.lt.0.000001) then
```

```
      x3=99999.
      go to 50
      endif
      a=1./((-log(omgw))**0.5)
      x3=a*sqrt(-log(omgw1))
50    continue
      do 40 j=1,36
      if(wel1.eq.wdat(j)) wel3=wdis(j)
      if(wel2.eq.wdat(j)) wel4=wdis(j)
40    continue
      if(tprob.gt.0.10) go to 9
      write(20,100) wel3,wel4,semiva,x1,x2,x3
      write(25,300) wel3,wel4,x1
9     continue
10    continue
11    continue
      close(10)
      close(20)
      close(25)
100   format(2a10,1x,4f8.5)
200   format(a9)
300   format(a9,',',a9,',',f7.5)
      end
```

VARIOPLOT PROGRAM

```
C       FORTRAN program uses data from variout to display
C       variograms from each well to all others. Data for scatter
C       plots put out to fort.(well number) files.
C
        character list*9,list1*9,well*9,wel2*9,wel3*9,wel4*9,char*3
        dimension list1(50),well(2000),wel2(2000),dist(2000),
       1 cdist(50,50),vario(2000),list(50),wel3(2000),wel4(2000),
       2 cvario(50,50)
        integer ijmax(50)
        open (unit=10,file='variout')
        open (unit=20,file='list')
        open (unit=30,file='list1')
        open (unit=40,file='distances')
        do 10 i=1,36
        read(30,100) list1(i)
        read(20,100) list(i)
 10     continue
        maxd=0
        do 20 k=1,999999
        read(40,200,end=21) well(k),wel2(k),dist(k)
        maxd=maxd+1
 20     continue
 21     continue
        maxv=0
        do 30 k=1,999999
        read(10,300,end=31) wel3(k),wel4(k),vario(k)
        maxv=maxv+1
 30     continue
 31     continue
        do 70 i=1,36
        ijmax(i)=0
        do 60 j=1,36
        if(i.eq.j) go to 60
        ii=i
        jj=j
        if(i.gt.j) then
        ii=j
        jj=i
        endif
        do 50 k=1,maxv
        if(wel3(k).eq.list1(ii).and.wel4(k).eq.list1(jj)) then
        do 40 l=1,maxd
        if(well(l).eq.list1(ii).and.wel2(l).eq.list1(jj)) then
        ijmax(i)=ijmax(i)+1
        cdist(i,ijmax(i))=dist(l)
        cvario(i,ijmax(i))=vario(k)
        endif
 40     continue
        endif
 50     continue
 60     continue
 70     continue
        do 90 i=1,36
        do 80 j=1,ijmax(i)-1
        do 75 k=j+1,ijmax(i)
        if(cdist(i,k).lt.cdist(i,j)) then
        buf=cdist(i,k)
        cdist(i,k)=cdist(i,j)
        cdist(i,j)=buf
        buf=cvario(i,k)
```

```
      cvario(i,k)=cvario(i,j)
      cvario(i,j)=buf
      endif
75    continue
80    continue
90    continue
      close(10)
      close(20)
      close(30)
      close(40)
      do 92 i=1,36
      open(unit=i+10)
      do 95 j=1,ijmax(i)
      write(i+10,*) cdist(i,j),cvario(i,j)
95    continue
      close(i+10)
92    continue
100   format(a9)
200   format(1x,a9,2x,a9,4x,f10.0)
300   format(1x,a9,1x,a9,f10.0,)
      end
```

EQSTATS PROGRAM

```
C      FORTRAN program uses data in variout to
C      calculate average differences between
C      spherical-exponential equations and between
C      exponential-Gaussian equations in calculated
C      CP radii. The same comparisons are expressed as
C      a percent. Results are sent to terminal screen.
C
       open(unit=10,file='variout')
       d12=0.
       d23=0.
       p12=0.
       p23=0.
       icount=0
       do 10 i=1,999999
       read(10,100,end=11,err=9) sphere,expon,gauss
       d12=d12+(sphere-expon)
       d23=d23+(expon-gauss)
       p12=p12+2.*(sphere-expon)/(sphere+expon)
       p23=p23+2.*(expon-gauss)/(expon+gauss)
       icount=icount+1
 9     continue
 10    continue
 11    continue
       close(10)
       d12=d12/float(icount)
       d23=d23/float(icount)
       p12=p12/float(icount)
       p23=p23/float(icount)
       write(*,*) d12,d23,p12,p23
 100   format(30x,3f8.0)
       end
```

VERIF PROGRAM

```
C      FORTRAN program uses data from files named
C      in user defined input file to compare with data
C      in the file table.2409. The name of the file that
C      contains the input file names and criteria is
C      specified by the user when executing. The
C      deviations between the data sets are analyzed and
C      the percent of data that is within the criteria
C      is output to the screen.
C
       DIMENSION H1A(8000),H2A(8000)
       INTEGER YR1,MO1,DA1,DATE1,YR2,MO2,DA2,DATE2
       CHARACTER JUNK*29,WEL1*15,WEL2*15,INPUT*9
 1     CONTINUE
       READ(*,*,END=500) INPUT,CRIT
       OPEN(UNIT=11,FILE='table.2409')
       OPEN(UNIT=12,FILE=INPUT)
       DO 5 I=1,2
       READ(11,*)
       READ(12,*)
 5     CONTINUE
       READ(11,300) JUNK,WEL1
       READ(12,300) JUNK,WEL2
       READ(11,*)
       READ(12,*)
       ICOUNT=0
       H1AVE=0.
       H2AVE=0.
       READ(11,*) YR1,MO1,DA1,H1
       READ(12,*) YR2,MO2,DA2,H2
 10    CONTINUE
       DATE1=10000*YR1+100*MO1+DA1
       DATE2=10000*YR2+100*MO2+DA2
       IF(DATE1.EQ.DATE2) GO TO 20
       IF(DATE1.GT.DATE2) THEN
       READ(12,*,END=99) YR2,MO2,DA2,H2
       GO TO 10
       ELSE
       READ(11,*,END=99) YR1,MO1,DA1,H1
       GO TO 10
       ENDIF
 20    CONTINUE
       IF(H1.GT.-1000.AND.H2.GT.-1000) THEN
       ICOUNT=ICOUNT+1
       H1A(ICOUNT)=H1
       H2A(ICOUNT)=H2
       H1AVE=H1AVE+H1
       H2AVE=H2AVE+H2
       ENDIF
       READ(11,*,END=99) YR1,MO1,DA1,H1
       READ(12,*,END=99) YR2,MO2,DA2,H2
       GO TO 10
 99    CONTINUE
       H1AVE=H1AVE/FLOAT(ICOUNT)
       H2AVE=H2AVE/FLOAT(ICOUNT)
       IC=0
       DO 30 I=1,ICOUNT
       IF(ABS(H1A(I)-H1AVE-H2A(I)+H2AVE).GT.CRIT)
      1 IC=IC+1
 30    CONTINUE
       WRITE(*,600) INPUT,CRIT,(FLOAT(ICOUNT)-FLOAT(IC))*100./
```

```
    1 FLOAT(ICOUNT)
      CLOSE(11)
      CLOSE(12)
      GO TO 1
500   CONTINUE
100   FORMAT(I4,I11,I13,F21.0)
200   FORMAT(I10,F16.2)
300   FORMAT(A29,A15)
400   FORMAT(2F16.5)
600   FORMAT(A10,1X,F5.2,1X,F4.1)
      END
```

CHK.COM SCRIPT

```
/* Unix script puts field data files into (well name).chk */
/* files without headers or missing data points. */

cat f291.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > f291.chk
cat g1213.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1213.chk
cat g1215.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1215.chk
cat g1220.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1220.chk
cat g1221.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1221.chk
cat g1222.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1222.chk
cat g1223.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1223.chk
cat g1224.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1224.chk
cat g1225.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1225.chk
cat g1226.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1226.chk
cat g1260.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1260.chk
cat g1262.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1262.chk
cat g1315.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1315.chk
cat g1316.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1316.chk
cat g1322.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1322.chk
cat g1472.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1472.chk
cat g1473.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1473.chk
cat g1636.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g1636.chk
cat g2030.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2030.chk
cat g2031.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2031.chk
cat g2032.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2032.chk
cat g2033.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2033.chk
cat g2034.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2034.chk
cat g2035.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2035.chk
cat g2147.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2147.chk
cat g2376.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2376.chk
cat g2395.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2395.chk
cat g2443.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2443.chk
cat g2444.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2444.chk
cat g2495.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g2495.chk
cat g561.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g561.chk
cat g616.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g616.chk
cat g617.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g617.chk
cat g820a.dat | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g820a.chk
cat g853.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > g853.chk
cat s329.dat   | grep -v 123456E20 | grep -v TIME | grep -v YEAR | grep -v “^$” | grep -v “\-\-” > s329.chk
```

AUTO.COM SCRIPT

```
/* Unix script executed in STATIT. Using data in */
/* (well name).chk files, the autocorrelation and */
/* autocovariance files are calculated. Output data */
/* goes to auto.dat. */
log auto.log
clear
read f291.chk into year,month,day,level /cases = 7000
delete year month day
acf level /nolist /scor /scov
let well = 1
delete level
write auto.dat all /noclose
clear
read g1213.chk into year,month,day,level /cases = 7000
delete year month day
acf level /nolist /scor /scov
let well = 1
delete level
write * all /noclose
clear
read g1215.chk into year,month,day,level /cases = 7000
delete year month day
acf level /nolist /scor /scov
let well = 1
delete level
write * all /noclose
clear
read g1220.chk into year,month,day,level /cases = 7000
delete year month day
acf level /nolist /scor /scov
let well = 1
delete level
write * all /noclose
clear
read g1221.chk into year,month,day,level /cases = 7000
delete year month day
acf level /nolist /scor /scov
let well = 1
delete level
write * all /noclose
clear

(CONTINUES FOR REMAINING WELLS)
```

AUTOVAR PROGRAM

```
C      FORTRAN program takes user input
C      autovariogram criteria and data in auto.dat.
C      Autovariograms are calculated for each well
C      (put out to fort.(well number) files) and
C      maximum sampling intervals (put out to autovar.out).

       character well*9
       write(*,*) 'enter autovariogram criteria'
       read(*,*) crit
       open(unit=90,file='auto.dat')
       open(unit=91,file='autovar.out')
       open(unit=92,file='list1')
       iflag=1
       icount=0
       iicoun=10
       autola=0.0
       do 10 i=1,999999
       read(90,100,end=20) corr,cov,icheck
       if(icheck.ne.0) then
       if (iflag.eq.0)  write(91,*)
      1 'For well ',well,' insufficient data'
       iflag=0
       icount=0
       iicoun=iicoun+1
       read (92,300) well
       write(iicoun,200) 0,0.0
       autola=0.0
       endif
       icount=icount+1
       autova=2*cov*(1/corr-1)
       write(iicoun,200) icount,autova
       if(autova.gt.crit.and.iflag.eq.0) then
       write(91,400) well,float(icount-1)+(crit-autola)/
      1 (autova-autola)
       iflag=1
       endif
       autola=autova
 10    continue
 20    continue
       close(90)
       close(91)
       close(92)
 100   format(2f8.0,i2)
 200   format(i5,1x,f7.3)
 300   format(a9)
 400   format(a9,1x,f5.2)
       end
```

APPENDIX 2. DERIVATION FOR LOG-NORMAL VARIATE

To describe the standard variate Z for a log-normal transformation, the following derivation was made. From Yevjevich (1982), the relation between the variance Ψ, the mean $\bar{\delta}$, and the variance of the log of the data Ψ_L is:

$$\Psi = \bar{\delta}^2 (e^{\Psi_L} - 1) \qquad \text{(A1)}$$

With rearrangement, this becomes:

$$\Psi_L = \ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right) \qquad \text{(A2)}$$

Also, from Yevjevich (1982), the relation of the above variables with the mean of the log of the data $\bar{\delta}_L$ is:

$$\bar{\delta} = e^{(\bar{\delta}_L + \Psi_L/2)} \qquad \text{(A3)}$$

With rearrangement, this becomes:

$$\bar{\delta}_L = \ln\bar{\delta} - \frac{\Psi_L}{2} \qquad \text{(A4)}$$

Combining equations A2 and A4 yields:

$$\bar{\delta}_L = \ln\bar{\delta} - \frac{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}{2} \qquad \text{(A5)}$$

The definition of the variate Z for a log-normal distribution is:

$$Z = \frac{(\ln\delta_i - \bar{\delta}_L)_{max}}{\sqrt{\Psi_L}} = \frac{(\ln\delta_{max}) - \bar{\delta}_L}{\sqrt{\Psi_L}} \qquad \text{(A6)}$$

where the subscript *max* indicates the values corresponding to $\delta - \bar{\delta} = \varepsilon$, the maximum allowable error in estimation. Substituting equations A2 and A5 into equation A6 yields:

$$Z = \frac{\ln\delta_{max} - \ln\bar{\delta} + \frac{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}{2}}{\sqrt{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}} \qquad \text{(A7)}$$

Because $\varepsilon = \delta_{max} - \bar{\delta}$, then $\delta_{max} = \varepsilon + \bar{\delta}$. Combining this expression and equation A7 and multiplying the numerator and denominator by two yields:

$$Z = \frac{2\ln(\varepsilon + \bar{\delta}) - 2\ln\bar{\delta} + \ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}{2\sqrt{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}} \qquad \text{(A8)}$$

Bringing the logarithmic terms together produces:

$$Z = \frac{2\ln\left(\frac{\varepsilon + \bar{\delta}}{\bar{\delta}}\right) + \ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}{2\sqrt{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}} \tag{A9}$$

Final rearrangement yields:

$$Z = \sqrt{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)}\left[\frac{\ln\left(\frac{\varepsilon + \bar{\delta}}{\bar{\delta}}\right)}{\ln\left(\frac{\Psi}{\bar{\delta}^2} + 1\right)} + \frac{1}{2}\right] \tag{A10}$$

This is the form used in equation 18 in the report.